

Université Bordeaux 1
Ecole doctorale mathématique et informatique

HABILITATION À DIRIGER DES RECHERCHES
Mathématiques appliquées

présentée par
Benoîte de Saporta

Contributions à l'estimation et au contrôle
de processus stochastiques

préparée chez Inria Bordeaux Sud-Ouest, Équipe CQFD
et à l'Institut de Mathématiques de Bordeaux

version du 11 mars 2013

Rapporteurs

Oswaldo Luis do Valle COSTA	Rapporteur	Escola Politécnica da Universidade de São Paulo
Jean-François DELMAS	Rapporteur	Ecole Nationale des Ponts et Chaussées
Gilles PAGÈS	Rapporteur	Université Pierre et Maris Curie

Table des matières

1	Articles relatifs au Chapitre 1	3
1.1	[BdSG09] B. BERCU, B. DE SAPORTA, AND A. GÉGOUT PETIT. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. <i>Electronic Journal of Probability</i> , scale=1.05, 14 :2492–2526, 2009.	3
1.2	[dSGP+11] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Parameters estimation for bifurcating autoregressive processes with missing data. <i>Electonic Journal of Statistics</i> , 5 :1313–1353, 2011.	39
1.3	[dSGP+12] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Asymmetry tests for bifurcating autoregressive processes with missing data. <i>Statistics & Probability Letters</i> , 82(7) :1439–1444, 2012.	81
1.4	[dSGP+13a] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Random coefficients bifurcating autoregressive processes. En révision pour ESAIM PS, arXiv :1205.3658, 2013.	88
1.5	[dSGP+13b] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Statistical study of asymmetry in cell lineage data. En révision pour Comp.l Statistics & Data Analysis, arXiv :1205.4840, 2013. . . .	125
2	Articles relatifs au Chapitre 2	151
2.1	[dSDG10] B. DE SAPORTA, F. DUFOUR, AND K. GONZALEZ. Numerical method for optimal stopping of piecewise deterministic Markov processes. <i>Ann. Appl. Probab.</i> , 20(5) :1607–1637, 2010.	151
2.2	[BBGP+13] C. BAYSSE, D. BIHANNIC, A. GÉGOUT PETIT, M. PRENAT, B. DE SAPORTA, AND J. SARACCO. Maintenance optimization of optronic equipment. In <i>Prognostics and System Health Management Conference</i> , Milano, Italy, 2013.	183
2.3	[dSDZ+12] B. DE SAPORTA, F. DUFOUR, H. ZHANG, AND C. ELEGBEDE. Optimal stopping for the predictive maintenance of a structure subject to corrosion. <i>journal of Risk and Reliability</i> , 226(2) :169–181, 2012.	190
2.4	[dSZ13] B. DE SAPORTA AND H. ZHANG. Predictive maintenance for the heated hold-up tank. <i>Reliability Engineering & System Safety</i> , 2013. To appear.	204
2.5	[BdSD13] A. BRANDEJSKY, B. DE SAPORTA, AND F. DUFOUR. Optimal stopping for partially observed piecewise-deterministic markov processes. <i>Stoch. Processes and their Applications</i> , 2013. To appear .	231
2.6	[dSD12] B. DE SAPORTA AND F. DUFOUR. Numerical method for impulse control of piecewise deterministic markov processes. <i>Automatica</i> , 48 :779–793, 2012.	273
3	Articles relatifs au Chapitre 3	289
3.1	[BdSD12a] A. BRANDEJSKY, B. DE SAPORTA, AND F. DUFOUR. Numerical method for expectations of piecewise-deterministic markov processes. <i>CAMCoS</i> , 7(1) :63–104, 2012.	289

3.2	[BdSD12b] A. BRANDEJSKY, B. DE SAPORTA, AND F. DUFOUR. Numerical methods for the exit time of a piecewise-deterministic markov process. <i>Advances in Applied Probability</i> , 44(1) :196–225, 2012.	335
4	Articles relatifs au Chapitre 4	366
4.1	[NML ⁺ 12] A. NÈGRE, O. MARCEAU, D. LANEUVILLE, H. ZHANG, B. DE SAPORTA, AND F. DUFOUR. Stochastic control for underwater optimal trajectories. In <i>IEEE Aerospace conference</i> , Big Sky, Montana, USA, 2012.	366
4.2	[ZdSD ⁺ 12] H. ZHANG, B. DE SAPORTA, F. DUFOUR, AND G. DELEUZE. Dynamic reliability : towards efficient simulation of the availability of a feedwater control system. In <i>NPIC-HMIT 2012</i> , San Diego, USA, 2012.	374

1 Articles relatifs au Chapitre 1

- 1.1 [BdSG09] B. BERCU, B. DE SAPORTA, AND A. GÉGOUT PETIT. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electronic Journal of Probability*, *scale=1.05*, 14 :2492–2526, 2009.

Asymptotic analysis for bifurcating autoregressive processes via a martingale approach

Bernard Bercu*

Benoîte de Saporta[†]

Anne Gégout-Petit[‡]

Abstract

We study the asymptotic behavior of the least squares estimators of the unknown parameters of general p th-order bifurcating autoregressive processes. Under very weak assumptions on the driven noise of the process, namely conditional pair-wise independence and suitable moment conditions, we establish the almost sure convergence of our estimators together with the quadratic strong law and the central limit theorem. All our analysis relies on non-standard asymptotic results for martingales.

Key words: bifurcating autoregressive process ; tree-indexed times series; martingales ; least squares estimation ; almost sure convergence ; quadratic strong law; central limit theorem.

AMS 2000 Subject Classification: Primary 60F15; Secondary: 60F05, 60G42.

Submitted to EJP on September 9, 2008, final version accepted November 2, 2009.

*Université de Bordeaux, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, bernard.bercu@math.u-bordeaux1.fr

[†]Université de Bordeaux, GREThA, CNRS, UMR 5113, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, saporta@math.u-bordeaux1.fr

[‡]Université de Bordeaux, IMB, CNRS, UMR 5251, and INRIA Bordeaux, team CQFD, France, anne.petit@u-bordeaux2.fr

1 Introduction

Bifurcating autoregressive (BAR) processes are an adaptation of autoregressive (AR) processes to binary tree structured data. They were first introduced by Cowan and Staudte [2] for cell lineage data, where each individual in one generation gives birth to two offspring in the next generation. Cell lineage data typically consist of observations of some quantitative characteristic of the cells over several generations of descendants from an initial cell. BAR processes take into account both inherited and environmental effects to explain the evolution of the quantitative characteristic under study.

More precisely, the original BAR process is defined as follows. The initial cell is labelled 1, and the two offspring of cell n are labelled $2n$ and $2n + 1$. Denote by X_n the quantitative characteristic of individual n . Then, the first-order BAR process is given, for all $n \geq 1$, by

$$\begin{cases} X_{2n} &= a + bX_n + \varepsilon_{2n}, \\ X_{2n+1} &= a + bX_n + \varepsilon_{2n+1}. \end{cases}$$

The noise sequence $(\varepsilon_{2n}, \varepsilon_{2n+1})$ represents environmental effects while a, b are unknown real parameters with $|b| < 1$. The driven noise $(\varepsilon_{2n}, \varepsilon_{2n+1})$ was originally supposed to be independent and identically distributed with normal distribution. However, two sister cells being in the same environment early in their lives, ε_{2n} and ε_{2n+1} are allowed to be correlated, inducing a correlation between sister cells distinct from the correlation inherited from their mother.

Several extensions of the model have been proposed. On the one hand, we refer the reader to Huggins and Basawa [10] and Basawa and Zhou [1; 15] for statistical inference on symmetric bifurcating processes. On the other hand, higher order processes, when not only the effects of the mother but also those of the grand-mother and higher order ancestors are taken into account, have been investigated by Huggins and Basawa [10]. Recently, an asymmetric model has been introduced by Guyon [5; 6] where only the effects of the mother are considered, but sister cells are allowed to have different conditional distributions. We can also mention a recent work of Delmas and Marsalle [3] dealing with a model of asymmetric bifurcating Markov chains on a Galton Watson tree instead of regular binary tree.

The purpose of this paper is to carry out a sharp analysis of the asymptotic properties of the least squares (LS) estimators of the unknown parameters of general asymmetric p th-order BAR processes. There are several results on statistical inference and asymptotic properties of estimators for BAR models in the literature. For maximum likelihood inference on small independent trees, see Huggins and Basawa [10]. For maximum likelihood inference on a single large tree, see Huggins [9] for the original BAR model, Huggins and Basawa [11] for higher order Gaussian BAR models, and Zhou and Basawa [15] for exponential first-order BAR processes. We also refer the reader to Zhou and Basawa [14] for the LS parameter estimation, and to Hwang, Basawa and Yeo [12] for the local asymptotic normality for BAR processes and related asymptotic inference. In all those papers, the process is supposed to be stationary. Consequently, X_n has a time-series representation involving an holomorphic function. In Guyon [5], the LS estimator is also investigated, but the process is not stationary, and the author makes intensive use of the tree structure and Markov chain theory. Our goal is to improve and extend the previous results of Guyon [5] via a martingale approach. As previously done by Basawa and Zhou [1; 14; 15] we shall make use of the strong law of large numbers [4] as well as the central limit theorem [7; 8] for martingales. It will allow

us to go further in the analysis of general p th-order BAR processes. We shall establish the almost sure convergence of the LS estimators together with the quadratic strong law and the central limit theorem.

The paper is organised as follows. Section 2 is devoted to the presentation of the asymmetric p th-order BAR process under study, while Section 3 deals with the LS estimators of the unknown parameters. In Section 4, we explain our strategy based on martingale theory. Our main results about the asymptotic properties of the LS estimators are given in Section 5. More precisely, we shall establish the almost sure convergence, the quadratic strong law (QSL) and the central limit theorem (CLT) for the LS estimators. The proof of our main results are detailed in Sections 6 to 10, the more technical ones being gathered in the appendices.

2 Bifurcating autoregressive processes

In all the sequel, let p be a non-zero integer. We consider the asymmetric $\text{BAR}(p)$ process given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} X_{2n} &= a_0 + \sum_{k=1}^p a_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n}, \\ X_{2n+1} &= b_0 + \sum_{k=1}^p b_k X_{[\frac{n}{2^{k-1}}]} + \varepsilon_{2n+1}, \end{cases} \quad (2.1)$$

where $[x]$ stands for the largest integer less than or equal to x . The initial states $\{X_k, 1 \leq k \leq 2^{p-1} - 1\}$ are the ancestors while $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is the driven noise of the process. The parameters (a_0, a_1, \dots, a_p) and (b_0, b_1, \dots, b_p) are unknown real numbers. The $\text{BAR}(p)$ process can be rewritten in the abbreviated vector form given, for all $n \geq 2^{p-1}$, by

$$\begin{cases} \mathbb{X}_{2n} &= A\mathbb{X}_n + \eta_{2n}, \\ \mathbb{X}_{2n+1} &= B\mathbb{X}_n + \eta_{2n+1}, \end{cases} \quad (2.2)$$

where the regression vector $\mathbb{X}_n = (X_n, X_{[\frac{n}{2}], \dots, X_{[\frac{n}{2^{p-1}}]})^t$, $\eta_{2n} = (a_0 + \varepsilon_{2n})e_1$, $\eta_{2n+1} = (b_0 + \varepsilon_{2n+1})e_1$ with $e_1 = (1, 0, \dots, 0)^t \in \mathbb{R}^p$. Moreover, A and B are the $p \times p$ companion matrices

$$A = \begin{pmatrix} a_1 & a_2 & \cdots & a_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad B = \begin{pmatrix} b_1 & b_2 & \cdots & b_p \\ 1 & 0 & \cdots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

This process is a direct generalization of the symmetric $\text{BAR}(p)$ process studied by Huggins, Basawa and Zhou [10; 14]. One can also observe that, in the particular case $p = 1$, it is the asymmetric BAR process studied by Guyon [5; 6]. In all the sequel, we shall assume that $\mathbb{E}[X_k^8] < \infty$ for all $1 \leq k \leq 2^{p-1} - 1$ and that matrices A and B satisfy the contracting property

$$\beta = \max\{\|A\|, \|B\|\} < 1,$$

where $\|A\| = \sup\{\|Au\|, u \in \mathbb{R}^p \text{ with } \|u\| = 1\}$.

As explained in the introduction, one can see this $\text{BAR}(p)$ process as a p th-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor, see Figure 1 for an illustration. For all $n \geq 1$, denote the n th generation by

$$\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}.$$

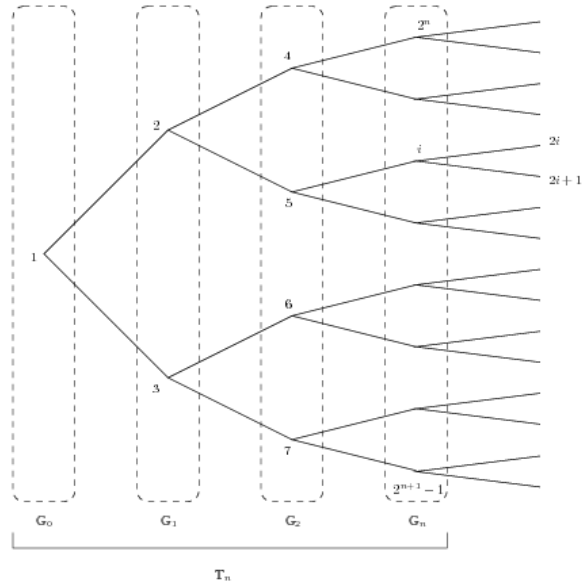


Figure 1: The tree associated with the bifurcating auto-regressive process.

In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the first ancestor. Let \mathbb{G}_{r_n} be the generation of individual n , which means that $r_n = \log_2(n)$. Recall that the two offspring of individual n are labelled $2n$ and $2n + 1$, or conversely, the mother of individual n is $\lfloor n/2 \rfloor$. More generally, the ancestors of individual n are $\lfloor n/2 \rfloor, \lfloor n/2^2 \rfloor, \dots, \lfloor n/2^{r_n} \rfloor$. Furthermore, denote by

$$\mathbb{T}_n = \bigcup_{k=0}^n \mathbb{G}_k$$

the sub-tree of all individuals from the original individual up to the n th generation. It is clear that the cardinality $|\mathbb{G}_n|$ of \mathbb{G}_n is 2^n while that of \mathbb{T}_n is $|\mathbb{T}_n| = 2^{n+1} - 1$. Finally, we denote by $\mathbb{T}_{n,p} = \{k \in \mathbb{T}_n, k \geq 2^p\}$ the sub-tree of all individuals up to the n th generation without \mathbb{T}_{p-1} . One can observe that, for all $n \geq 1$, $\mathbb{T}_{n,0} = \mathbb{T}_n$ and, for all $p \geq 1$, $\mathbb{T}_{p,p} = \mathbb{G}_p$.

3 Least-squares estimation

The BAR(p) process (2.1) can be rewritten, for all $n \geq 2^{p-1}$, in the matrix form

$$Z_n = \theta^t Y_n + V_n \tag{3.1}$$

where

$$Z_n = \begin{pmatrix} X_{2n} \\ X_{2n+1} \end{pmatrix}, \quad Y_n = \begin{pmatrix} 1 \\ \mathbb{X}_n \end{pmatrix}, \quad V_n = \begin{pmatrix} \varepsilon_{2n} \\ \varepsilon_{2n+1} \end{pmatrix},$$

and the $(p + 1) \times 2$ matrix parameter θ is given by

$$\theta = \begin{pmatrix} a_0 & b_0 \\ a_1 & b_1 \\ \vdots & \vdots \\ a_p & b_p \end{pmatrix}.$$

Our goal is to estimate θ from the observation of all individuals up to the n th generation that is the complete sub-tree \mathbb{T}_n . Each new generation \mathbb{G}_n contains half the global available information. Consequently, we shall show that observing the whole tree \mathbb{T}_n or only generation \mathbb{G}_n is almost the same. We propose to make use of the standard LS estimator $\hat{\theta}_n$ which minimizes

$$\Delta_n(\theta) = \frac{1}{2} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|Z_k - \theta^t Y_k\|^2.$$

Consequently, we obviously have for all $n \geq p$

$$\hat{\theta}_n = S_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1, p-1}} Y_k Z_k^t, \quad (3.2)$$

where the $(p + 1) \times (p + 1)$ matrix S_n is defined as

$$S_n = \sum_{k \in \mathbb{T}_{n, p-1}} Y_k Y_k^t = \sum_{k \in \mathbb{T}_{n, p-1}} \begin{pmatrix} 1 & \mathbb{X}_k^t \\ \mathbb{X}_k & \mathbb{X}_k \mathbb{X}_k^t \end{pmatrix}.$$

In the special case where $p = 1$, S_n simply reduces to

$$S_n = \sum_{k \in \mathbb{T}_n} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

In order to avoid useless invertibility assumption, we shall assume, without loss of generality, that for all $n \geq p - 1$, S_n is invertible. Otherwise, we only have to add the identity matrix I_{p+1} to S_n . In all what follows, we shall make a slight abuse of notation by identifying θ as well as $\hat{\theta}_n$ to

$$\text{vec}(\theta) = \begin{pmatrix} a_0 \\ \vdots \\ a_p \\ b_0 \\ \vdots \\ b_p \end{pmatrix} \quad \text{and} \quad \text{vec}(\hat{\theta}_n) = \begin{pmatrix} \hat{a}_{0,n} \\ \vdots \\ \hat{a}_{p,n} \\ \hat{b}_{0,n} \\ \vdots \\ \hat{b}_{p,n} \end{pmatrix}.$$

The reason for this change will be explained in Section 4. Hence, we readily deduce from (3.2) that

$$\begin{aligned} \hat{\theta}_n &= (I_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \text{vec}(Y_k Z_k^t) \\ &= (I_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} X_{2k} \\ X_k \mathbb{X}_{2k} \\ X_{2k+1} \\ X_k \mathbb{X}_{2k+1} \end{pmatrix}, \end{aligned}$$

where \otimes stands for the matrix Kronecker product. Consequently, it follows from (3.1) that

$$\begin{aligned}\widehat{\theta}_n - \theta &= (\mathbf{I}_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \text{vec} (Y_k V_k^t) \\ &= (\mathbf{I}_2 \otimes S_{n-1}^{-1}) \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix}.\end{aligned}\tag{3.3}$$

Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the BAR(p) process, which means that \mathcal{F}_n is the σ -algebra generated by all individuals up to the n th generation, $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$. In all the sequel, we shall make use of the five following moment hypotheses.

(H.1) One can find $\sigma^2 > 0$ such that, for all $n \geq p - 1$ and for all $k \in \mathbb{G}_{n+1}$, ε_k belongs to L^2 with

$$\mathbb{E}[\varepsilon_k | \mathcal{F}_n] = 0 \quad \text{and} \quad \mathbb{E}[\varepsilon_k^2 | \mathcal{F}_n] = \sigma^2 \quad \text{a.s.}$$

(H.2) It exists $|\rho| < \sigma^2$ such that, for all $n \geq p - 1$ and for all different $k, l \in \mathbb{G}_{n+1}$ with $[k/2] = [l/2]$,

$$\mathbb{E}[\varepsilon_k \varepsilon_l | \mathcal{F}_n] = \rho \quad \text{a.s.}$$

Otherwise, ε_k and ε_l are conditionally independent given \mathcal{F}_n .

(H.3) For all $n \geq p - 1$ and for all $k \in \mathbb{G}_{n+1}$, ε_k belongs to L^4 and

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

(H.4) One can find $\tau^4 > 0$ such that, for all $n \geq p - 1$ and for all $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] = \tau^4 \quad \text{a.s.}$$

and, for $\nu^2 < \tau^4$ and for all different $k, l \in \mathbb{G}_{n+1}$ with $[k/2] = [l/2]$

$$\mathbb{E}[\varepsilon_{2k}^2 \varepsilon_{2k+1}^2 | \mathcal{F}_n] = \nu^2 \quad \text{a.s.}$$

(H.5) For all $n \geq p - 1$ and for all $k \in \mathbb{G}_{n+1}$, ε_k belongs to L^8 with

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[\varepsilon_k^8 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

Remark 3.1. In contrast with [14], one can observe that we do not assume that $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is a sequence of independent and identically distributed bi-variate random vectors. The price to pay for giving up this iid assumption is higher moments, namely assumptions **(H.3)** and **(H.5)**. Indeed we need them to make use of the strong law of large numbers and the central limit theorem for martingales. However, we do not require any normality assumption on $(\varepsilon_{2n}, \varepsilon_{2n+1})$. Consequently, our assumptions are much weaker than the existing ones in previous literature.

We now turn to the estimation of the parameters σ^2 and ρ . On the one hand, we propose to estimate the conditional variance σ^2 by

$$\widehat{\sigma}_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k\|^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2) \quad (3.4)$$

where for all $n \geq p - 1$ and for all $k \in \mathbb{G}_n$, $\widehat{V}_k^t = (\widehat{\varepsilon}_{2k}, \widehat{\varepsilon}_{2k+1})$ with

$$\begin{cases} \widehat{\varepsilon}_{2k} &= X_{2k} - \widehat{a}_{0,n} - \sum_{i=1}^p \widehat{a}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor}, \\ \widehat{\varepsilon}_{2k+1} &= X_{2k+1} - \widehat{b}_{0,n} - \sum_{i=1}^p \widehat{b}_{i,n} X_{\lfloor \frac{k}{2^{i-1}} \rfloor}. \end{cases}$$

One can observe that, on the above equations, we make use of only the past observations for the estimation of the parameters. This will be crucial in the asymptotic analysis. On the other hand, we estimate the conditional covariance ρ by

$$\widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1}. \quad (3.5)$$

4 Martingale approach

In order to establish all the asymptotic properties of our estimators, we shall make use of a martingale approach. It allows us to impose a very smooth restriction on the driven noise (ε_n) compared with the previous results in the literature. As a matter of fact, we only assume suitable moment conditions on (ε_n) and that $(\varepsilon_{2n}, \varepsilon_{2n+1})$ are conditionally independent, while it is assumed in [14] that $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is a sequence of independent identically distributed random vectors. For all $n \geq p$, denote

$$M_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k} \\ \varepsilon_{2k} \mathbb{X}_k \\ \varepsilon_{2k+1} \\ \varepsilon_{2k+1} \mathbb{X}_k \end{pmatrix} \in \mathbb{R}^{2(p+1)}.$$

Let $\Sigma_n = I_2 \otimes S_n$, and note that $\Sigma_n^{-1} = I_2 \otimes S_n^{-1}$. For all $n \geq p$, we can thus rewrite (3.3) as

$$\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n. \quad (4.1)$$

The key point of our approach is that (M_n) is a martingale. Most of all the asymptotic results for martingales were established for vector-valued martingales. That is the reason why we have chosen to make use of vector notation in Section 3. In order to show that (M_n) is a martingale adapted to the filtration $\mathbb{F} = (\mathcal{F}_n)$, we rewrite it in a compact form. Let $\Psi_n = I_2 \otimes \Phi_n$, where Φ_n is the rectangular matrix of dimension $(p+1) \times \delta_n$, with $\delta_n = 2^n$, given by

$$\Phi_n = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \mathbb{X}_{2^n} & \mathbb{X}_{2^{n+1}} & \cdots & \mathbb{X}_{2^{n+1}-1} \end{pmatrix}.$$

It contains the individuals of generations \mathbb{G}_{n-p+1} up to \mathbb{G}_n and is also the collection of all Y_k , $k \in \mathbb{G}_n$. Let ξ_n be the random vector of dimension δ_n

$$\xi_n = \begin{pmatrix} \varepsilon_{2^n} \\ \varepsilon_{2^{n+2}} \\ \vdots \\ \varepsilon_{2^{n+1}-2} \\ \varepsilon_{2^{n+1}} \\ \varepsilon_{2^{n+3}} \\ \vdots \\ \varepsilon_{2^{n+1}-1} \end{pmatrix}.$$

The vector ξ_n gathers the noise variables of generation \mathbb{G}_n . The special ordering separating odd and even indices is tailor-made so that M_n can be written as

$$M_n = \sum_{k=p}^n \Psi_{k-1} \xi_k.$$

By the same token, one can observe that

$$S_n = \sum_{k=p-1}^n \Phi_k \Phi_k^t \quad \text{and} \quad \Sigma_n = \sum_{k=p-1}^n \Psi_k \Psi_k^t.$$

Under **(H.1)** and **(H.2)**, we clearly have for all $n \geq 0$, $\mathbb{E}[\xi_{n+1} | \mathcal{F}_n] = 0$ and Ψ_n is \mathcal{F}_n -measurable. In addition, it is not hard to see that for all $n \geq 0$, $\mathbb{E}[\xi_{n+1} \xi_{n+1}^t | \mathcal{F}_n] = \Gamma \otimes I_{\delta_n}$ where Γ is the covariance matrix associated with $(\varepsilon_{2^n}, \varepsilon_{2^{n+1}})$

$$\Gamma = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix}.$$

We shall also prove that (M_n) is a square integrable martingale. Its increasing process is given for all $n \geq p + 1$ by

$$\langle M \rangle_n = \sum_{k=p-1}^{n-1} \Psi_k (\Gamma \otimes I_{\delta_k}) \Psi_k^t = \Gamma \otimes \sum_{k=p-1}^{n-1} \Phi_k \Phi_k^t = \Gamma \otimes S_{n-1}.$$

It is necessary to establish the convergence of S_n , properly normalized, in order to prove the asymptotic results for the $\text{BAR}(p)$ estimators $\hat{\theta}_n$, $\hat{\sigma}_n^2$ and $\hat{\rho}_n$. One can observe that the sizes of Ψ_n and ξ_n are not fixed and double at each generation. This is why we have to adapt the proof of vector-valued martingale convergence given in [4] to our framework.

5 Main results

We now state our main results, first on the martingale (M_n) and then on our estimators.

Proposition 5.1. Assume that (ε_n) satisfies **(H.1)** to **(H.3)**. Then, we have

$$\lim_{n \rightarrow \infty} \frac{S_n}{|\mathbb{T}_n|} = L \quad \text{a.s.} \quad (5.1)$$

where L is a positive definite matrix specified in Section 7.

This result is the keystone of our asymptotic analysis. It enables us to prove sharp asymptotic properties for (M_n) .

Theorem 5.1. Assume that (ε_n) satisfies **(H.1)** to **(H.3)**. Then, we have

$$M_n^t \Sigma_{n-1}^{-1} M_n = \mathcal{O}(n) \quad \text{a.s.} \quad (5.2)$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.3)$$

Moreover, if (ε_n) satisfies **(H.4)** and **(H.5)**, we have the central limit theorem

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L). \quad (5.4)$$

From the asymptotic properties of (M_n) , we deduce the asymptotic behavior of our estimators. Our first result deals with the almost sure asymptotic properties of the LS estimator $\widehat{\theta}_n$.

Theorem 5.2. Assume that (ε_n) satisfies **(H.1)** to **(H.3)**. Then, $\widehat{\theta}_n$ converges almost surely to θ with the rate of convergence

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}|}{|\mathbb{T}_{n-1}|}\right) \quad \text{a.s.} \quad (5.5)$$

In addition, we also have the quadratic strong law

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n |\mathbb{T}_{k-1}| (\widehat{\theta}_k - \theta)^t \Lambda (\widehat{\theta}_k - \theta) = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.6)$$

where $\Lambda = I_2 \otimes L$.

Our second result is devoted to the almost sure asymptotic properties of the variance and covariance estimators $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n$. Let

$$\sigma_n^2 = \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} (\varepsilon_{2k}^2 + \varepsilon_{2k+1}^2) \quad \text{and} \quad \rho_n = \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p}} \varepsilon_{2k} \varepsilon_{2k+1}.$$

Theorem 5.3. Assume that (ε_n) satisfies **(H.1)** to **(H.3)**. Then, $\widehat{\sigma}_n^2$ converges almost surely to σ^2 . More precisely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1,p}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2 = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (5.7)$$

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) = 2(p+1)\sigma^2 \quad a.s. \quad (5.8)$$

In addition, $\widehat{\rho}_n$ converges almost surely to ρ

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1,p}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = (p+1)\rho \quad a.s. \quad (5.9)$$

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\rho}_n - \rho_n) = 2(p+1)\rho \quad a.s. \quad (5.10)$$

Our third result concerns the asymptotic normality for all our estimators $\widehat{\theta}_n$, $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n$.

Theorem 5.4. Assume that (ε_n) satisfies (H.1) to (H.5). Then, we have the central limit theorem

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L^{-1}). \quad (5.11)$$

In addition, we also have

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\tau^4 - 2\sigma^4 + \nu^2}{2}\right) \quad (5.12)$$

and

$$\sqrt{|\mathbb{T}_{n-1}|}(\widehat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 - \rho^2). \quad (5.13)$$

The rest of the paper is dedicated to the proof of our main results. We start by giving laws of large numbers for the noise sequence (ε_n) in Section 6. In Section 7, we give the proof of Proposition 5.1. Sections 8, 9 and 10 are devoted to the proofs of Theorems 5.2, 5.3 and 5.4, respectively. The more technical proofs, including that of Theorem 5.1, are postponed to the Appendices.

6 Laws of large numbers for the noise sequence

We first need to establish strong laws of large numbers for the noise sequence (ε_n) . These results will be useful in all the sequel. We will extensively use the strong law of large numbers for locally square integrable real martingales given in Theorem 1.3.15 of [4].

Lemma 6.1. Assume that (ε_n) satisfies (H.1) and (H.2). Then

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k = 0 \quad a.s. \quad (6.1)$$

In addition, if (H.3) holds, we also have

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2 = \sigma^2 \quad a.s. \quad (6.2)$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k} \varepsilon_{2k+1} = \rho \quad a.s. \quad (6.3)$$

Proof: On the one hand, let

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \varepsilon_i.$$

We have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_{n+1}} \varepsilon_k.$$

Hence, it follows from (H.1) and (H.2) that (P_n) is a square integrable real martingale with increasing process

$$\langle P \rangle_n = (\sigma^2 + \rho) \sum_{k=p}^n |\mathbb{G}_k| = (\sigma^2 + \rho)(|\mathbb{T}_n| - |\mathbb{T}_{p-1}|).$$

Consequently, we deduce from Theorem 1.3.15 of [4] that $P_n = o(\langle P \rangle_n)$ a.s. which implies (6.1). On the other hand, denote

$$Q_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} e_i,$$

where $e_n = \varepsilon_n^2 - \sigma^2$. We have

$$\Delta Q_{n+1} = Q_{n+1} - Q_n = \frac{1}{|\mathbb{G}_{n+1}|} \sum_{k \in \mathbb{G}_{n+1}} e_k.$$

First of all, it follows from (H.1) that for all $k \in \mathbb{G}_{n+1}$, $\mathbb{E}[e_k | \mathcal{F}_n] = 0$ a.s. In addition, for all different $k, l \in \mathbb{G}_{n+1}$ with $[k/2] \neq [l/2]$,

$$\mathbb{E}[e_k e_l | \mathcal{F}_n] = 0 \quad \text{a.s.}$$

thanks to the conditional independence given by (H.2). Furthermore, we readily deduce from (H.3) that

$$\sup_{n \geq p-1} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[e_k^2 | \mathcal{F}_n] < \infty \quad \text{a.s.}$$

Therefore, (Q_n) is a square integrable real martingale with increasing process

$$\begin{aligned} \langle Q \rangle_n &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] \sum_{j=p}^n \frac{1}{|\mathbb{G}_j|} \quad \text{a.s.} \\ &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] \sum_{j=p}^n \left(\frac{1}{2}\right)^j \quad \text{a.s.} \\ &\leq 2 \sup_{p-1 \leq k \leq n-1} \sup_{i \in \mathbb{G}_{k+1}} \mathbb{E}[e_i^2 | \mathcal{F}_k] < \infty \quad \text{a.s.} \end{aligned}$$

Consequently, we obtain from the strong law of large numbers for martingales that (Q_n) converges almost surely. Finally, as $(|\mathbb{G}_n|)$ is a positive real sequence which increases to infinity, we find from Lemma A.1 in Appendix A that

$$\sum_{k=p}^n \sum_{i \in \mathbb{G}_k} e_i = o(|\mathbb{G}_n|) \quad \text{a.s.}$$

leading to

$$\sum_{k=p}^n \sum_{i \in \mathbb{G}_k} e_i = o(|\mathbb{T}_n|) \quad \text{a.s.}$$

as $|\mathbb{T}_n| - 1 = 2|\mathbb{G}_n|$, which implies (6.2). We also establish (6.3) in a similar way. As a matter of fact, let

$$R_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_{k-1}|} \sum_{i \in \mathbb{G}_{k-1}} (\varepsilon_{2i} \varepsilon_{2i+1} - \rho).$$

Then, (R_n) is a square integrable real martingale which converges almost surely, leading to (6.3). \square

Remark 6.2. Note that via Lemma A.2

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k} &= 0, & \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1} &= 0 & \text{a.s.} \\ \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k}^2 &= \sigma^2, & \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1}^2 &= \sigma^2 & \text{a.s.} \end{aligned}$$

In fact, each new generation contains half the global available information, observing the whole tree \mathbb{T}_n or only generation \mathbb{G}_n is essentially the same.

For the CLT, we will also need the convergence of higher moments of the driven noise (ε_n) .

Lemma 6.3. Assume that (ε_n) satisfies (H.1) to (H.5). Then, we have

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^4 = \tau^4 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1,p-1}} \varepsilon_{2k}^2 \varepsilon_{2k+1}^2 = \nu^2 \quad \text{a.s.}$$

Proof : The proof is left to the reader as it follows essentially the same lines as the proof of Lemma 6.1 using the square integrable real martingales

$$Q_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} (\varepsilon_i^4 - \tau^4)$$

and

$$R_n = \sum_{k=p}^n \frac{1}{|\mathbb{G}_{k-1}|} \sum_{i \in \mathbb{G}_{k-1}} (\varepsilon_{2i}^2 \varepsilon_{2i+1}^2 - \nu^2).$$

Remark 6.4. Note that again via Lemma A.2

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k}^4 = \tau^4 \quad \text{and} \quad \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_{2k+1}^4 = \tau^4 \quad \text{a.s.}$$

7 Proof of Proposition 5.1

Proposition 5.1 is a direct application of the two following lemmas which provide two strong laws of large numbers for the sequence of random vectors (\mathbb{X}_n) .

Lemma 7.1. *Assume that (ε_n) satisfies (H.1) and (H.2). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k = \lambda = \bar{a}(I_p - \bar{A})^{-1}e_1 \quad a.s. \quad (7.1)$$

where $\bar{a} = (a_0 + b_0)/2$ and \bar{A} is the mean of the companion matrices

$$\bar{A} = \frac{1}{2}(A + B).$$

Lemma 7.2. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p}} \mathbb{X}_k \mathbb{X}_k^t = \ell, \quad a.s. \quad (7.2)$$

where the matrix ℓ is the unique solution of the equation

$$\ell = T + \frac{1}{2}(A\ell A^t + B\ell B^t)$$

$$T = (\sigma^2 + \bar{a}^2)e_1 e_1^t + \frac{1}{2}(a_0(A\lambda e_1^t + e_1 \lambda^t A^t) + b_0(B\lambda e_1^t + e_1 \lambda^t B^t))$$

with $\bar{a}^2 = (a_0^2 + b_0^2)/2$.

Proof : The proofs are given in Appendix A. □

Remark 7.3. *We shall see in Appendix A that*

$$\ell = \sum_{k=0}^{\infty} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} CTC^t$$

where the notation $\{A;B\}^k$ means the set of all products of A and B with exactly k terms. For example, we have $\{A;B\}^0 = \{I_p\}$, $\{A;B\}^1 = \{A, B\}$, $\{A;B\}^2 = \{A^2, AB, BA, B^2\}$ and so on. The cardinality of $\{A;B\}^k$ is obviously 2^k .

Remark 7.4. *One can observe that in the special case $p = 1$,*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} X_k &= \frac{\bar{a}}{1 - \bar{b}} \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} X_k^2 &= \frac{\bar{a}^2 + \sigma^2 + 2\lambda \bar{a} \bar{b}}{1 - \bar{b}^2} \quad a.s. \end{aligned}$$

where

$$\bar{a} \bar{b} = \frac{a_0 a_1 + b_0 b_1}{2}, \quad \bar{b} = \frac{a_1 + b_1}{2}, \quad \bar{b}^2 = \frac{a_1^2 + b_1^2}{2}.$$

8 Proof of Theorems 5.1 and 5.2

Theorem 5.2 is a consequence of Theorem 5.1. The first result of Theorem 5.1 is a strong law of large numbers for the martingale (M_n) . We already mentioned that the standard strong law is useless here. This is due to the fact that the dimension of the random vector ξ_n grows exponentially fast as 2^n . Consequently, we are led to propose a new strong law of large numbers for (M_n) , adapted to our framework.

Proof of result (5.2) of Theorem 5.1: For all $n \geq p$, let $\mathcal{V}_n = M_n^t \Sigma_{n-1}^{-1} M_n$ where we recall that $\Sigma_n = I_2 \otimes S_n$, so that $\Sigma_n^{-1} = I_2 \otimes S_n^{-1}$. First of all, we have

$$\begin{aligned} \mathcal{V}_{n+1} &= M_{n+1}^t \Sigma_n^{-1} M_{n+1} = (M_n + \Delta M_{n+1})^t \Sigma_n^{-1} (M_n + \Delta M_{n+1}), \\ &= M_n^t \Sigma_n^{-1} M_n + 2M_n^t \Sigma_n^{-1} \Delta M_{n+1} + \Delta M_{n+1}^t \Sigma_n^{-1} \Delta M_{n+1}, \\ &= \mathcal{V}_n - M_n^t (\Sigma_{n-1}^{-1} - \Sigma_n^{-1}) M_n + 2M_n^t \Sigma_n^{-1} \Delta M_{n+1} + \Delta M_{n+1}^t \Sigma_n^{-1} \Delta M_{n+1}. \end{aligned}$$

By summing over this identity, we obtain the main decomposition

$$\mathcal{V}_{n+1} + \mathcal{A}_n = \mathcal{V}_p + \mathcal{B}_{n+1} + \mathcal{W}_{n+1}, \quad (8.1)$$

where

$$\begin{aligned} \mathcal{A}_n &= \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k, \\ \mathcal{B}_{n+1} &= 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Delta M_{k+1} \quad \text{and} \quad \mathcal{W}_{n+1} = \sum_{k=p}^n \Delta M_{k+1}^t \Sigma_k^{-1} \Delta M_{k+1}. \end{aligned}$$

The asymptotic behavior of the left-hand side of (8.1) is as follows.

Lemma 8.1. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{V}_{n+1} + \mathcal{A}_n}{n} = (p+1)\sigma^2 \quad \text{a.s.} \quad (8.2)$$

Proof: The proof is given in Appendix B. It relies on the Riccati equation associated to (S_n) and the strong law of large numbers for (\mathcal{W}_n) . \square

Since (\mathcal{V}_n) and (\mathcal{A}_n) are two sequences of positive real numbers, we infer from Lemma 8.1 that $\mathcal{V}_{n+1} = \mathcal{O}(n)$ a.s. which ends the proof of (5.2). \square

Proof of result (5.5) of Theorem 5.2: It clearly follows from (4.1) that

$$\mathcal{V}_n = (\widehat{\theta}_n - \theta)^t \Sigma_{n-1} (\widehat{\theta}_n - \theta).$$

Consequently, the asymptotic behavior of $\widehat{\theta}_n - \theta$ is clearly related to the one of \mathcal{V}_n . More precisely, we can deduce from convergence (5.1) that

$$\lim_{n \rightarrow \infty} \frac{\lambda_{\min}(\Sigma_n)}{|\mathbb{T}_n|} = \lambda_{\min}(\Lambda) > 0 \quad \text{a.s.}$$

since L as well as $\Lambda = I_2 \otimes L$ are definite positive matrices. Here $\lambda_{\min}(\Lambda)$ stands for the smallest eigenvalue of the matrix Λ . Therefore, as

$$\|\widehat{\theta}_n - \theta\|^2 \leq \frac{\mathcal{V}_n}{\lambda_{\min}(\Sigma_{n-1})},$$

we use (5.2) to conclude that

$$\|\widehat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{n}{|\mathbb{T}_{n-1}|}\right) = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}|}{|\mathbb{T}_{n-1}|}\right) \quad \text{a.s.}$$

which completes the proof of (5.5). □

We now turn to the proof of the quadratic strong law. To this end, we need a sharper estimate of the asymptotic behavior of (\mathcal{V}_n) .

Lemma 8.2. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have for all $\delta > 1/2$,*

$$\|M_n\|^2 = o(|\mathbb{T}_{n-1}|n^\delta) \quad \text{a.s.} \quad (8.3)$$

Proof: The proof is given in Appendix C. □

A direct application of Lemma 8.2 ensures that $\mathcal{V}_n = o(n^\delta)$ a.s. for all $\delta > 1/2$. Hence, Lemma 8.1 immediately leads to the following result.

Corollary 8.3. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{A}_n}{n} = (p+1)\sigma^2 \quad \text{a.s.} \quad (8.4)$$

Proof of result (5.3) of Theorem 5.1: First of all, \mathcal{A}_n may be rewritten as

$$\mathcal{A}_n = \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k = \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1/2} \Delta_k \Sigma_{k-1}^{-1/2} M_k$$

where $\Delta_n = I_{2(p+1)} - \Sigma_{n-1}^{1/2} \Sigma_n^{-1} \Sigma_{n-1}^{1/2}$. In addition, via Proposition 5.1

$$\lim_{n \rightarrow \infty} \frac{\Sigma_n}{|\mathbb{T}_n|} = \Lambda \quad \text{a.s.} \quad (8.5)$$

which implies that

$$\lim_{n \rightarrow \infty} \Delta_n = \frac{1}{2} I_{2(p+1)} \quad \text{a.s.} \quad (8.6)$$

Furthermore, it follows from Corollary 8.3 that $\mathcal{A}_n = \mathcal{O}(n)$ a.s. Hence, we deduce from (8.5) and (8.6) that

$$\frac{\mathcal{A}_n}{n} = \left(\frac{1}{2n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k \right) + o(1) \quad \text{a.s.} \quad (8.7)$$

and convergence (5.3) directly follows from Corollary 8.3. \square

We are now in position to prove the QSL.

Proof of result (5.6) of Theorem 5.2: The QSL is a direct consequence of (5.3) together with the fact that $\widehat{\theta}_n - \theta = \Sigma_{n-1}^{-1} M_n$. Indeed, we have

$$\begin{aligned} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k &= \frac{1}{n} \sum_{k=p}^n (\widehat{\theta}_k - \theta)^t \Sigma_{k-1} (\widehat{\theta}_k - \theta) \\ &= \frac{1}{n} \sum_{k=p}^n |\mathbb{T}_{k-1}| (\widehat{\theta}_k - \theta)^t \frac{\Sigma_{k-1}}{|\mathbb{T}_{k-1}|} (\widehat{\theta}_k - \theta) \\ &= \frac{1}{n} \sum_{k=p}^n |\mathbb{T}_{k-1}| (\widehat{\theta}_k - \theta)^t \Lambda (\widehat{\theta}_k - \theta) + o(1) \quad \text{a.s.} \end{aligned}$$

which completes the proof of Theorem 5.2. \square

9 Proof of Theorem 5.3

The almost sure convergence of $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n$ is strongly related to that of $\widehat{V}_n - V_n$.

Proof of result (5.7) of Theorem 5.3: We need to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = 2(p+1)\sigma^2 \quad \text{a.s.} \quad (9.1)$$

Once again, we are searching for a link between the sum of $\|\widehat{V}_n - V_n\|$ and the processes (\mathcal{A}_n) and (\mathcal{Y}_n) whose convergence properties were previously investigated. For all $n \geq p$, we have

$$\begin{aligned} \sum_{k \in \mathbb{G}_n} \|\widehat{V}_k - V_k\|^2 &= \sum_{k \in \mathbb{G}_n} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2, \\ &= (\widehat{\theta}_n - \theta)^t \Psi_n \Psi_n^t (\widehat{\theta}_n - \theta), \\ &= M_n^t \Sigma_{n-1}^{-1} \Psi_n \Psi_n^t \Sigma_{n-1}^{-1} M_n, \\ &= M_n^t \Sigma_{n-1}^{-1/2} \Delta_n \Sigma_{n-1}^{-1/2} M_n, \end{aligned}$$

where

$$\Delta_n = \Sigma_{n-1}^{-1/2} \Psi_n \Psi_n^t \Sigma_{n-1}^{-1/2} = \Sigma_{n-1}^{-1/2} (\Sigma_n - \Sigma_{n-1}) \Sigma_{n-1}^{-1/2}.$$

Now, we can deduce from convergence (8.5) that

$$\lim_{n \rightarrow \infty} \Delta_n = I_{2(p+1)} \quad \text{a.s.}$$

which implies that

$$\sum_{k \in \mathbb{G}_n} \|\widehat{V}_k - V_k\|^2 = M_n^t \Sigma_{n-1}^{-1} M_n (1 + o(1)) \quad \text{a.s.}$$

Therefore, we can conclude via convergence (5.3) that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=p}^n M_k^t \Sigma_{k-1}^{-1} M_k = 2(p+1)\sigma^2 \quad \text{a.s.}$$

Proof of result (5.8) of Theorem 5.3: First of all,

$$\begin{aligned} \widehat{\sigma}_n^2 - \sigma_n^2 &= \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\|\widehat{V}_k\|^2 - \|V_k\|^2), \\ &= \frac{1}{2|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\|\widehat{V}_k - V_k\|^2 + 2(\widehat{V}_k - V_k)^t V_k). \end{aligned}$$

Set

$$P_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t V_k = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} (\widehat{V}_i - V_i)^t V_i.$$

We clearly have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_n} (\widehat{V}_k - V_k)^t V_k.$$

One can observe that for all $k \in \mathbb{G}_n$, $\widehat{V}_k - V_k = (I_2 \otimes Y_k)^t (\theta - \widehat{\theta}_n)$ which implies that $\widehat{V}_k - V_k$ is \mathcal{F}_n -measurable. Consequently, (P_n) is a real martingale transform. Hence, we can deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] together with (9.1) that

$$P_n = o\left(\sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2\right) = o(n) \quad \text{a.s.}$$

It ensures once again via convergence (9.1) that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 = 2(p+1)\sigma^2 \quad \text{a.s.}$$

We now turn to the study of the covariance estimator $\widehat{\rho}_n$. We have

$$\begin{aligned} \widehat{\rho}_n - \rho_n &= \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1} - \varepsilon_{2k} \varepsilon_{2k+1}), \\ &= \frac{1}{|\mathbb{T}_{n-1}|} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) + \frac{1}{|\mathbb{T}_{n-1}|} Q_n, \end{aligned}$$

where

$$Q_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \varepsilon_{2k+1} + (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \varepsilon_{2k} = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t J_2 V_k$$

with

$$J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Moreover, one can observe that $J_2 \Gamma J_2 = \Gamma$. Hence, as before, (Q_n) is a real martingale transform satisfying

$$Q_n = o \left(\sum_{k \in \mathbb{T}_{n-1, p-1}} \|\widehat{V}_k - V_k\|^2 \right) = o(n) \quad \text{a.s.}$$

We will see in Appendix D that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = (p+1)\rho \quad \text{a.s.} \quad (9.2)$$

Finally, we find from (9.2) that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n|}{n} (\widehat{\rho}_n - \rho_n) = 2(p+1)\rho \quad \text{a.s.}$$

which completes the proof of Theorem 5.3. \square

10 Proof of Theorem 5.4

In order to prove the CLT for the BAR(p) estimators, we will use the central limit theorem for martingale difference sequences given in Propositions 7.8 and 7.9 of Hamilton [8].

Proposition 10.1. *Assume that (W_n) is a vector martingale difference sequence satisfying*

(a) *For all $n \geq 1$, $\mathbb{E}[W_n W_n^t] = \Omega_n$ where Ω_n is a positive definite matrix and*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \Omega_k = \Omega$$

where Ω is also a positive definite matrix.

(b) *For all $n \geq 1$ and for all i, j, k, l , $\mathbb{E}[W_{in} W_{jn} W_{kn} W_{ln}] < \infty$ where W_{in} is the i th element of the vector W_n .*

(c)

$$\frac{1}{n} \sum_{k=1}^n W_k W_k^t \xrightarrow{\mathcal{P}} \Omega.$$

Then, we have the central limit theorem

$$\frac{1}{\sqrt{n}} \sum_{k=1}^n W_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Omega).$$

We wish to point out that for $\text{BAR}(p)$ processes, it seems impossible to make use of the standard CLT for martingales. This is due to the fact that Lindeberg's condition is not satisfied in our framework. Moreover, as the size of (ξ_n) doubles at each generation, it is also impossible to check condition (c). To overcome this problem, we simply change the filtration. Instead of using the generation-wise filtration, we will use the sister pair-wise one. Let

$$\mathcal{G}_n = \sigma\{X_1, (X_{2k}, X_{2k+1}), 1 \leq k \leq n\}$$

be the σ -algebra generated by all pairs of individuals up to the offspring of individual n . Hence $(\varepsilon_{2n}, \varepsilon_{2n+1})$ is \mathcal{G}_n -measurable. Note that \mathcal{G}_n is also the σ -algebra generated by, on the one hand, all the past generations up to that of individual n , i.e. the r_n th generation, and, on the other hand, all pairs of the $(r_n + 1)$ th generation with ancestors less than or equal to n . In short,

$$\mathcal{G}_n = \sigma\left(\mathcal{F}_{r_n} \cup \{(X_{2k}, X_{2k+1}), k \in \mathbb{G}_{r_n}, k \leq n\}\right).$$

Therefore, (H.2) implies that the processes $(\varepsilon_{2n}, \mathbb{X}_n \varepsilon_{2n}, \varepsilon_{2n+1}, \mathbb{X}_n \varepsilon_{2n+1})^t$, $(\varepsilon_{2n}^2 + \varepsilon_{2n+1}^2 - 2\sigma^2)$ and $(\varepsilon_{2n} \varepsilon_{2n+1} - \rho)$ are \mathcal{G}_n -martingales.

Proof of result (5.4) of Theorem 5.1: First, recall that $Y_n = (1, \mathbb{X}_n)^t$. We apply Propositions 10.1 to the \mathcal{G}_n -martingale difference sequence (D_n) given by

$$D_n = \text{vec}(Y_n V_n^t) = \begin{pmatrix} \varepsilon_{2n} \\ \mathbb{X}_n \varepsilon_{2n} \\ \varepsilon_{2n+1} \\ \mathbb{X}_n \varepsilon_{2n+1} \end{pmatrix}.$$

We clearly have

$$D_n D_n^t = \begin{pmatrix} \varepsilon_{2n}^2 & \varepsilon_{2n} \varepsilon_{2n+1} \\ \varepsilon_{2n+1} \varepsilon_{2n} & \varepsilon_{2n+1}^2 \end{pmatrix} \otimes Y_n Y_n^t.$$

Hence, it follows from (H.1) and (H.2) that

$$\mathbb{E}[D_n D_n^t] = \Gamma \otimes \mathbb{E}[Y_n Y_n^t].$$

Moreover, we can show by a slight change in the proof of Lemmas 7.1 and 7.2 that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n-1, p-1}} \mathbb{E}[D_k D_k^t] = \Gamma \otimes \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \mathbb{E}[S_n] = \Gamma \otimes L,$$

which is positive definite, so that condition (a) holds. Condition (b) also clearly holds under (H.3). We now turn to condition (c). We have

$$\sum_{k \in \mathbb{T}_{n-1, p-1}} D_k D_k^t = \Gamma \otimes S_n + R_n$$

where

$$R_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \begin{pmatrix} \varepsilon_{2k}^2 - \sigma^2 & \varepsilon_{2k} \varepsilon_{2k+1} - \rho \\ \varepsilon_{2k+1} \varepsilon_{2k} - \rho & \varepsilon_{2k+1}^2 - \sigma^2 \end{pmatrix} \otimes Y_k Y_k^t.$$

Under (H.1) to (H.5), we can show that (R_n) is a martingale transform. Moreover, we can prove that $R_n = o(n)$ a.s. using Lemma A.6 and similar calculations as in Appendix B where a more complicated martingale transform (K_n) is studied. Consequently, condition (c) also holds and we can conclude that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1, p-1}} D_k = \frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma \otimes L). \quad (10.1)$$

Proof of result (5.11) of Theorem 5.4: We deduce from (4.1) that

$$\sqrt{|\mathbb{T}_{n-1}|}(\hat{\theta}_n - \theta) = |\mathbb{T}_{n-1}| \Sigma_{n-1}^{-1} \frac{M_n}{\sqrt{|\mathbb{T}_{n-1}|}}.$$

Hence, (5.11) directly follows from (5.4) and convergence (8.5) together with Slutsky's Lemma. \square

Proof of results (5.12) and (5.13) of Theorem 5.4: On the one hand, we apply Propositions 10.1 to the \mathcal{G}_n -martingale difference sequence (v_n) defined by

$$v_n = \varepsilon_{2n}^2 + \varepsilon_{2n+1}^2 - 2\sigma^2.$$

Under (H.4), one has $\mathbb{E}[v_n^2] = 2\tau^4 - 4\sigma^4 + 2v^2$ which ensures that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} \mathbb{E}[v_k^2] = 2\tau^4 - 4\sigma^4 + 2v^2 > 0.$$

Hence, condition (a) holds. Once again, condition (b) clearly holds under (H.5), and Lemma 6.3 together with Remark 6.4 imply condition (c),

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} v_k^2 = 2\tau^4 - 4\sigma^4 + 2v^2 \quad \text{a.s.}$$

Therefore, we obtain that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1, p-1}} v_k = 2\sqrt{|\mathbb{T}_{n-1}|}(\sigma_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 2\tau^4 - 4\sigma^4 + 2v^2). \quad (10.2)$$

Furthermore, we infer from (5.8) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}|}(\hat{\sigma}_n^2 - \sigma_n^2) = 0 \quad \text{a.s.} \quad (10.3)$$

Finally, (10.2) and (10.3) imply (5.12). On the other hand, we apply again Proposition 10.1 to the \mathcal{G}_n -martingale difference sequence (w_n) given by

$$w_n = \varepsilon_{2n}\varepsilon_{2n+1} - \rho.$$

Under (H.4), one has $\mathbb{E}[w_n^2] = v^2 - \rho^2$ which implies that condition (a) holds since

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n, p-1}} \mathbb{E}[w_k^2] = v^2 - \rho^2 > 0.$$

Once again, condition (b) clearly holds under **(H.5)**, and Lemmas 6.1 and 6.3 yield condition (c),

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_{n,p-1}} w_k^2 = v^2 - \rho^2 \quad \text{a.s.}$$

Consequently, we obtain that

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}|}} \sum_{k \in \mathbb{T}_{n-1,p-1}} w_k = \sqrt{|\mathbb{T}_{n-1}|}(\rho_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, v^2 - \rho^2). \quad (10.4)$$

Furthermore, we infer from (5.10) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}|}(\widehat{\rho}_n - \rho_n) = 0 \quad \text{a.s.} \quad (10.5)$$

Finally, (5.13) follows from (10.4) and (10.5) which completes the proof of Theorem 5.4. \square

Appendices

A Laws of large numbers for the BAR process

We start with some technical Lemmas we make repeatedly use of, the well-known Kronecker's Lemma given in Lemma 1.3.14 of [4] together with some related results.

Lemma A.1. *Let (α_n) be a sequence of positive real numbers increasing to infinity. In addition, let (x_n) be a sequence of real numbers such that*

$$\sum_{n=0}^{\infty} \frac{|x_n|}{\alpha_n} < +\infty.$$

Then, one has

$$\lim_{n \rightarrow \infty} \frac{1}{\alpha_n} \sum_{k=0}^n x_k = 0.$$

Lemma A.2. *Let (x_n) be a sequence of real numbers. Then,*

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k = x \iff \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k = x. \quad (\text{A.1})$$

Proof: First of all, recall that $|\mathbb{T}_n| = 2^{n+1} - 1$ and $|\mathbb{G}_n| = 2^n$. Assume that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k = x.$$

We have the decomposition,

$$\sum_{k \in \mathbb{T}_n} x_k = \sum_{k \in \mathbb{T}_{n-1}} x_k + \sum_{k \in \mathbb{G}_n} x_k.$$

Consequently,

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k &= \lim_{n \rightarrow \infty} \frac{2}{|\mathbb{T}_n| + 1} \sum_{k \in \mathbb{T}_n} x_k - \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_{n-1}| + 1} \sum_{k \in \mathbb{T}_{n-1}} x_k, \\ &= 2x - x = x.\end{aligned}$$

Conversely, suppose that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} x_k = x.$$

A direct application of Toeplitz Lemma given in Lemma 2.2.13 of [4]) yields

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k \in \mathbb{T}_n} x_k &= \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k=0}^n \sum_{i \in \mathbb{G}_k} x_i, \\ &= \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_n|} \sum_{k=0}^n 2^k \frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} x_i = x.\end{aligned}$$

Lemma A.3. Let (A_n) be a sequence of real-valued matrices such that $\sum_{n=0}^{\infty} \|A_n\| < \infty$ and

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n A_k = A.$$

In addition, let (X_n) be a sequence of real-valued vectors which converges to a limiting value X . Then,

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n A_{n-k} X_k = AX. \quad (\text{A.2})$$

Proof: For all $n \geq 0$, let

$$U_n = \sum_{k=0}^n A_{n-k} X_k.$$

We clearly have for all integer n_0 with $1 \leq n_0 < n$,

$$\begin{aligned}\|U_n - AX\| &= \left\| \sum_{k=0}^n A_{n-k} X_k - \sum_{k=0}^n A_k X - \sum_{k=n+1}^{\infty} A_k X \right\|, \\ &\leq \sum_{k=0}^n \|A_{n-k}\| \|X_k - X\| + \sum_{k=n+1}^{\infty} \|A_k\| \|X\|, \\ &\leq \sum_{k=0}^{n_0} \|A_{n-k}\| \|X_k - X\| + \sum_{k=n_0+1}^n \|A_{n-k}\| \|X_k - X\| + \sum_{k=n+1}^{\infty} \|A_k\| \|X\|.\end{aligned}$$

We assume that (X_n) converges to a limiting value X . Consequently, we can choose n_0 such that for all $k > n_0$, $\|X_k - X\| < \varepsilon$. Moreover, one can find $M > 0$ such that for all $k \geq 0$, $\|X_k - X\| \leq M$ and $\|X\| \leq M$. Therefore, we obtain that

$$\|U_n - AX\| \leq (n_0 + 1)M \sup_{k \geq n-n_0} \|A_k\| + \varepsilon \sum_{k=n_0+1}^n \|A_{n-k}\| + M \sum_{k=n+1}^{\infty} \|A_k\|.$$

On the one hand

$$\sup_{k \geq n-n_0} \|A_k\| \quad \text{and} \quad \sum_{k=n+1}^{\infty} \|A_k\|$$

both converge to 0 as n tends to infinity. On the other hand,

$$\sum_{k=n_0+1}^n \|A_{n-k}\| \leq \sum_{n=0}^{\infty} \|A_n\| < \infty.$$

Consequently, $\|U_n - AX\|$ goes to 0 as n goes to infinity, as expected. \square

Lemma A.4. *Let (T_n) be a convergent sequence of real-valued matrices with limiting value T . Then,*

$$\lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T_{n-k} C^t = \ell$$

where the matrix

$$\ell = \sum_{k=0}^{\infty} \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C T C^t$$

is the unique solution of the equation

$$\ell = T + \frac{1}{2}(A\ell A^t + B\ell B^t). \quad (\text{A.3})$$

Proof: First of all, recall that $\beta = \max\{\|A\|, \|B\|\} < 1$. The cardinality of $\{A;B\}^k$ is obviously 2^k . Consequently, if

$$U_n = \sum_{k=0}^n \frac{1}{2^k} \sum_{C \in \{A;B\}^k} C(T_{n-k} - T)C^t,$$

it is not hard to see that

$$\|U_n\| \leq \sum_{k=0}^n \frac{1}{2^k} \times 2^k \beta^{2k} \|T_{n-k} - T\| = \sum_{k=0}^n \beta^{2(n-k)} \|T_k - T\|.$$

Hence, (U_n) converges to zero which completes the proof of Lemma A.4. \square

We now return to the BAR process. We first need an estimate of the sum of the $\|\mathbb{X}_n\|^2$ before being able to investigate the limits.

Lemma A.5. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\sum_{k \in \mathbb{T}_{n,p}^+} \|\mathbb{X}_k\|^2 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.} \quad (\text{A.4})$$

Proof: In all the sequel, for all $n \geq 2^{p-1}$, denote $A_{2n} = A$ and $A_{2n+1} = B$. It follows from a recursive application of relation (2.2) that for all $n \geq 2^{p-1}$

$$\mathbb{X}_n = \left(\prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} + \sum_{k=0}^{r_n-p} \left(\prod_{i=0}^{k-1} A_{\lfloor \frac{n}{2^i} \rfloor} \right) \eta_{\lfloor \frac{n}{2^k} \rfloor} \quad (\text{A.5})$$

with the convention that an empty product equals 1. Then, we can deduce from Cauchy-Schwarz inequality that for all $n \geq 2^{p-1}$

$$\begin{aligned} \left\| \mathbb{X}_n - \left(\prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} \right\|^2 &= \left\| \sum_{k=0}^{r_n-p} \left(\prod_{i=0}^{k-1} A_{\lfloor \frac{n}{2^i} \rfloor} \right) \eta_{\lfloor \frac{n}{2^k} \rfloor} \right\|^2 \\ &\leq \left(\sum_{k=0}^{r_n-p} \left(\prod_{i=0}^{k-1} \|A_{\lfloor \frac{n}{2^i} \rfloor}\| \right) \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\| \right)^2 \\ &\leq \left(\sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\| \right)^2 \\ &\leq \left(\sum_{k=0}^{r_n-p} \beta^k \right) \left(\sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right) \\ &\leq \frac{1}{1-\beta} \left(\sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right). \end{aligned}$$

Hence, we obtain that for all $n \geq 2^p$,

$$\begin{aligned} \|\mathbb{X}_n\|^2 &= \left\| \mathbb{X}_n - \left(\prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} + \left(\prod_{k=0}^{r_n-p} A_{\lfloor \frac{n}{2^k} \rfloor} \right) \mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor} \right\|^2 \\ &\leq \frac{2}{1-\beta} \left(\sum_{k=0}^{r_n-p} \beta^k \|\eta_{\lfloor \frac{n}{2^k} \rfloor}\|^2 \right) + 2\beta^{2(r_n-p+1)} \|\mathbb{X}_{\lfloor \frac{n}{2^{r_n-p+1}} \rfloor}\|^2. \end{aligned}$$

Denote $\alpha = \max\{|a_0|, |b_0|\}$ and $\bar{X}_1 = \max\{\|X_k\|, k \leq 2^{p-1}\}$. Summing up over the sub-tree $\mathbb{T}_{n,p}$, we find that

$$\begin{aligned} \sum_{k \in \mathbb{T}_{n,p}} \|\mathbb{X}_k\|^2 &\leq \sum_{k \in \mathbb{T}_{n,p}} \frac{2}{1-\beta} \left(\sum_{i=0}^{r_k-p} \beta^i \|\eta_{\lfloor \frac{k}{2^i} \rfloor}\|^2 \right) + \sum_{k \in \mathbb{T}_{n,p}} 2\beta^{2(r_k-p+1)} \|\mathbb{X}_{\lfloor \frac{k}{2^{r_k-p+1}} \rfloor}\|^2 \\ &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i (\alpha^2 + \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2) + \sum_{k \in \mathbb{T}_{n,p}} 2\beta^{2(r_k-p+1)} \|X_{\lfloor \frac{k}{2^{r_k-p+1}} \rfloor}\|^2 \\ &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2 + \frac{4\alpha^2}{1-\beta} \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \\ &\quad + 2\bar{X}_1^2 \sum_{k \in \mathbb{T}_{n,p}} \beta^{2(r_k-p+1)}, \\ &\leq \frac{4P_n}{1-\beta} + \frac{4\alpha^2 Q_n}{1-\beta} + 2\bar{X}_1^2 R_n, \end{aligned} \quad (\text{A.6})$$

where

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^2, \quad Q_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{r_k-p} \beta^i, \quad R_n = \sum_{k \in \mathbb{T}_{n,p}} \beta^{2(r_k-p+1)}.$$

The last two terms of (A.6) are readily evaluated by splitting the sums generation-wise. As a matter of fact,

$$Q_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \frac{1-\beta^k}{1-\beta} \leq \frac{1}{(1-\beta)} \sum_{k=p}^n 2^k = \mathcal{O}(|\mathbb{T}_n|), \quad (\text{A.7})$$

and

$$R_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_k} \beta^{k-p+1} \leq \sum_{k=p}^n (2\beta)^k = \mathcal{O}(|\mathbb{T}_n|). \quad (\text{A.8})$$

It remains to control the first term P_n . One can observe that ε_k appears in P_n as many times as it has descendants up to the n th generation, and its multiplicative factor for its i th generation descendant is $(2\beta)^i$. Hence, one has

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \sum_{i=0}^{n-r_k} (2\beta)^i \varepsilon_k^2.$$

The evaluation of P_n depends on the value of $0 < \beta < 1$. On the one hand, if $\beta = 1/2$, P_n reduces to

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} (n+1-r_k) \varepsilon_k^2 = \sum_{k=p}^n (n+1-k) \sum_{i \in \mathbb{G}_k} \varepsilon_i^2.$$

Hence,

$$\frac{P_n}{|\mathbb{T}_n| + 1} = \sum_{k=p}^n \left(\frac{(n+1-k)}{2^{n+1-k}} \right) \left(\frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} \varepsilon_i^2 \right).$$

However, it follows from Remark 6.2 that

$$\lim_{n \rightarrow +\infty} \frac{1}{|\mathbb{G}_n|} \sum_{k \in \mathbb{G}_n} \varepsilon_k^2 = \sigma^2 \quad \text{a.s.}$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n \frac{k}{2^k} = 2.$$

Consequently, we infer from Lemma A.3 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{|\mathbb{T}_n|} = 2\sigma^2 \quad \text{a.s.} \quad (\text{A.9})$$

On the other hand, if $\beta \neq 1/2$, we have

$$P_n = \sum_{k \in \mathbb{T}_{n,p}} \frac{1-(2\beta)^{n-r_k+1}}{1-2\beta} \varepsilon_k^2 = \frac{1}{1-2\beta} \sum_{k=p}^n (1-(2\beta)^{n-k+1}) \sum_{i \in \mathbb{G}_k} \varepsilon_i^2.$$

Thus,

$$\frac{P_n}{|\mathbb{T}_n| + 1} = \frac{1}{1 - 2\beta} \sum_{k=p}^n \left(\left(\frac{1}{2} \right)^{n-k+1} - \beta^{n-k+1} \right) \left(\frac{1}{|\mathbb{G}_k|} \sum_{i \in \mathbb{G}_k} \varepsilon_i^2 \right).$$

Furthermore,

$$\lim_{n \rightarrow \infty} \frac{1}{1 - 2\beta} \sum_{k=1}^n \left(\left(\frac{1}{2} \right)^k - \beta^k \right) = \frac{1}{1 - \beta}.$$

As before, we deduce from Lemma A.3 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{|\mathbb{T}_n|} = \frac{\sigma^2}{1 - \beta}. \quad \text{a.s.} \quad (\text{A.10})$$

Finally, Lemma A.5 follows from the conjunction of (A.6), (A.7), (A.8) together with (A.9) and (A.10). \square

Proof of Lemma 7.1 : First of all, denote

$$H_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \quad \text{and} \quad P_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k,$$

As $|\mathbb{T}_n| = 2^{n+1} - 1$, we obtain from Equation (2.2) the recursive relation

$$\begin{aligned} H_n &= H_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \left(A_k \mathbb{X}_{\lfloor \frac{k}{2} \rfloor} + \eta_k \right), \\ &= H_{p-1} + 2\bar{A}H_{n-1} + 2\bar{a}(2^n - 2^{p-1})e_1 + P_n e_1 \end{aligned} \quad (\text{A.11})$$

where $e_1 = (1, 0, \dots, 0)^t \in \mathbb{R}^p$, $\bar{a} = (a_0 + b_0)/2$ and the matrix

$$\bar{A} = \frac{A+B}{2}.$$

By induction, we deduce from (A.11) that

$$\begin{aligned} \frac{H_n}{2^{n+1}} &= \frac{H_{p-1}}{2^{n+1}} + \bar{A} \frac{H_{n-1}}{2^n} + \bar{a} \left(\frac{2^n - 2^{p-1}}{2^n} \right) e_1 + \frac{P_n}{2^{n+1}} e_1, \\ &= (\bar{A})^{n-p+1} \frac{H_{p-1}}{2^p} + \sum_{k=p}^n (\bar{A})^{n-k} \left(\frac{H_{p-1}}{2^{k+1}} + \bar{a} \left(\frac{2^k - 2^{p-1}}{2^k} \right) e_1 + \frac{P_k}{2^{k+1}} e_1 \right). \end{aligned}$$

We have already seen via convergence (6.1) of Lemma 6.1 that

$$\lim_{n \rightarrow +\infty} \frac{P_n}{2^{n+1}} = 0 \quad \text{a.s.}$$

Finally, as $\|\bar{A}\| < 1$,

$$\sum_{n=0}^{\infty} \|(\bar{A})^n\| < \infty \quad \text{and} \quad (\text{I}_p - \bar{A})^{-1} = \sum_{n=0}^{\infty} (\bar{A})^n,$$

it follows from Lemma A.3 that

$$\lim_{n \rightarrow \infty} \frac{H_n}{2^{n+1}} = \bar{a}(I_p - \bar{A})^{-1}e_1 \quad \text{a.s.}$$

which ends the proof of Lemma 7.1. \square

Proof of Lemma 7.2 : We shall proceed as in the proof of Lemma 7.1 and use the same notation. Let

$$K_n = \sum_{k \in \mathbb{T}_{n,p-1}} \mathbb{X}_k \mathbb{X}_k^t \quad \text{and} \quad L_n = \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2.$$

We infer again from (2.2) that

$$\begin{aligned} K_n &= K_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \left(A_k \mathbb{X}_{[\frac{k}{2}]} + \eta_k \right) \left(A_k \mathbb{X}_{[\frac{k}{2}]} + \eta_k \right)^t \\ &= K_{p-1} + \sum_{k \in \mathbb{T}_{n,p}} \varepsilon_k^2 e_1 e_1^t + \sum_{k \in \mathbb{T}_{n-1,p-1}} \left(A \mathbb{X}_k \mathbb{X}_k^t A^t + B \mathbb{X}_k \mathbb{X}_k^t B^t \right) \\ &\quad + \sum_{k \in \mathbb{T}_{n-1,p-1}} \left((a_0 + \varepsilon_{2k}) \mathbb{U}_k(A) + (b_0 + \varepsilon_{2k+1}) \mathbb{U}_k(B) + 2(\bar{a} + \zeta_{2k}) e_1 e_1^t \right) \end{aligned}$$

where $\mathbb{U}_k(A) = A \mathbb{X}_k e_1^t + e_1 \mathbb{X}_k^t A^t$ and $\mathbb{U}_k(B) = B \mathbb{X}_k e_1^t + e_1 \mathbb{X}_k^t B^t$. In addition, $\bar{a}^2 = (a_0^2 + b_0^2)/2$ and $\zeta_{2k} = (a_0 \varepsilon_{2k} + b_0 \varepsilon_{2k+1})$. Therefore, we obtain that

$$\frac{K_n}{2^{n+1}} = \frac{1}{2} \left(A \frac{K_{n-1}}{2^n} A^t + B \frac{K_{n-1}}{2^n} B^t \right) + T_n \quad (\text{A.12})$$

where

$$\begin{aligned} T_n &= \left(\frac{L_n}{2^{n+1}} + \bar{a}^2 \left(\frac{2^n - 2^{p-1}}{2^n} \right) + \frac{1}{2^n} \sum_{k \in \mathbb{T}_{n-1,p-1}} \zeta_{2k} \right) e_1 e_1^t \\ &\quad + \frac{1}{2} \left(a_0 \left(A \frac{H_{n-1}}{2^n} e_1^t + e_1 \frac{H_{n-1}^t}{2^n} A^t \right) + b_0 \left(B \frac{H_{n-1}}{2^n} e_1^t + e_1 \frac{H_{n-1}^t}{2^n} B^t \right) \right) \\ &\quad + \frac{1}{2^{n+1}} \sum_{k \in \mathbb{T}_{n-1,p-1}} \left(\varepsilon_{2k} \mathbb{U}_k(A) + \varepsilon_{2k+1} \mathbb{U}_k(B) \right). \end{aligned}$$

The two first results (6.1) and (6.2) of Lemma 6.1 together with Remark 6.2 and Lemma A.2 readily imply that

$$\lim_{n \rightarrow +\infty} \frac{L_n}{2^{n+1}} = \sigma^2 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow +\infty} \frac{1}{2^n} \sum_{k \in \mathbb{T}_{n-1,p-1}} \zeta_{2k} = 0 \quad \text{a.s.}$$

In addition, Lemma 7.1 gives

$$\lim_{n \rightarrow +\infty} \frac{H_{n-1}}{2^n} = \lambda \quad \text{a.s.}$$

Furthermore, denote

$$U_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} \left(\varepsilon_{2k} \mathbb{U}_k(A) + \varepsilon_{2k+1} \mathbb{U}_k(B) \right).$$

For all $u \in \mathbb{R}^p$, let $U_n(u) = u^t U_n u$. The sequence $(U_n(u))$ is a real martingale transform. Moreover, it follows from Lemma A.5 that

$$\sum_{k \in \mathbb{T}_{n-1, p-1}} \left| u^t \mathbb{U}_k(A) u \right|^2 + \left| u^t \mathbb{U}_k(B) u \right|^2 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.}$$

Consequently, we deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] that $U_n(u) = o(|\mathbb{T}_n|)$ a.s. for all $u \in \mathbb{R}^p$ which leads to $U_n = o(|\mathbb{T}_n|)$ a.s. Therefore, we obtain that (T_n) converges a.s. to T given by

$$T = (\sigma^2 + \overline{a^2}) e_1 e_1^t + \frac{1}{2} \left(A \lambda a_0 e_1^t + a_0 e_1 \lambda^t A^t + B \lambda b_0 e_1^t + b_0 e_1 \lambda^t B^t \right).$$

Finally, iteration of the recursive relation (A.12) yields

$$\frac{K_n}{2^{n+1}} = \frac{1}{2^{n-p+1}} \sum_{C \in \{A; B\}^{n-p+1}} C \frac{K_{p-1}}{2^p} C^t + \sum_{k=0}^{n-p} \frac{1}{2^k} \sum_{C \in \{A; B\}^k} C T_{n-k} C^t.$$

On the one hand, the first term on the right-hand side converges a.s. to zero as its norm is bounded $\beta^{2(n-p+1)} \|K_{p-1}\| / 2^p$. On the other hand, thanks to Lemma A.4, the second term on the right-hand side converges to ℓ given by (A.3), which completes the proof of Lemma 7.2. \square

We now state a convergence result for the sum of $\|\mathbb{X}_n\|^4$ which will be useful for the CLT.

Lemma A.6. *Assume that (ε_n) satisfies (H.1) to (H.5). Then, we have*

$$\sum_{k \in \mathbb{T}_{n, p}} \|\mathbb{X}_k\|^4 = \mathcal{O}(|\mathbb{T}_n|) \quad \text{a.s.} \quad (\text{A.13})$$

Proof : The proof is almost exactly the same as that of Lemma A.5. Instead of Equation (A.6), we have

$$\sum_{k \in \mathbb{T}_{n, p}} \|\mathbb{X}_k\|^4 \leq \frac{64 P_n}{(1-\beta)^3} + \frac{64 \alpha^4 Q_n}{(1-\beta)^3} + 8 \overline{X}_1^4 R_n$$

where

$$P_n = \sum_{k \in \mathbb{T}_{n, p}} \sum_{i=0}^{r_k-p} \beta^i \varepsilon_{\lfloor \frac{k}{2^i} \rfloor}^4, \quad Q_n = \sum_{k \in \mathbb{T}_{n, p}} \sum_{i=0}^{r_k-p} \beta^i, \quad R_n = \sum_{k \in \mathbb{T}_{n, p}} \beta^{4(r_k-p+1)}.$$

We already saw that $Q_n = \mathcal{O}(|\mathbb{T}_n|)$. In addition, it is not hard to see that $R_n = \mathcal{O}(|\mathbb{T}_n|)$. Therefore, we only need a sharper estimate for u_n . Via the same lines as in the proof of Lemma A.5 together with the sharper results of Lemma 6.3, we can show that $P_n = \mathcal{O}(|\mathbb{T}_n|)$ a.s. which leads to (A.13). \square

B On the quadratic strong law

We start with an auxiliary lemma closely related to the Riccati Equation for the inverse of the matrix S_n .

Lemma B.1. *Let h_n and l_n be the two following symmetric square matrices of order δ_n*

$$h_n = \Phi_n^t S_n^{-1} \Phi_n \quad \text{and} \quad l_n = \Phi_n^t S_{n-1}^{-1} \Phi_n.$$

Then, the inverse of S_n may be recursively calculated as

$$S_n^{-1} = S_{n-1}^{-1} - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} \Phi_n^t S_{n-1}^{-1}. \quad (\text{B.1})$$

In addition, we also have $(\mathbf{I}_{\delta_n} - h_n)(\mathbf{I}_{\delta_n} + l_n) = \mathbf{I}_{\delta_n}$.

Remark B.2. *If $f_n = \Psi_n^t \Sigma_n^{-1} \Psi_n$, it follows from Lemma B.1 that*

$$\Sigma_n^{-1} = \Sigma_{n-1}^{-1} - \Sigma_{n-1}^{-1} \Psi_n (\mathbf{I}_{2\delta_n} - f_n) \Psi_n^t \Sigma_{n-1}^{-1}. \quad (\text{B.2})$$

Proof : As $S_n = S_{n-1} + \Phi_n \Phi_n^t$, relation (B.1) immediately follows from Riccati Equation given e.g. in [4] page 96. By multiplying both side of (B.1) by Φ_n , we obtain

$$\begin{aligned} S_n^{-1} \Phi_n &= S_{n-1}^{-1} \Phi_n - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} l_n, \\ &= S_{n-1}^{-1} \Phi_n - S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1} (\mathbf{I}_{\delta_n} + l_n - \mathbf{I}_{\delta_n}), \\ &= S_{n-1}^{-1} \Phi_n (\mathbf{I}_{\delta_n} + l_n)^{-1}. \end{aligned}$$

Consequently, multiplying this time on the left by Φ_n^t , we obtain that

$$\begin{aligned} h_n &= l_n (\mathbf{I}_{\delta_n} + l_n)^{-1} = (l_n + \mathbf{I}_{\delta_n} - \mathbf{I}_{\delta_n}) (\mathbf{I}_{\delta_n} + l_n)^{-1}, \\ &= \mathbf{I}_{\delta_n} - (\mathbf{I}_{\delta_n} + l_n)^{-1} \end{aligned}$$

leading to $(\mathbf{I}_{\delta_n} - h_n)(\mathbf{I}_{\delta_n} + l_n) = \mathbf{I}_{\delta_n}$. □

In order to establish the quadratic strong law for (M_n) , we are going to study separately the asymptotic behaviour of (\mathcal{W}_n) and (\mathcal{B}_n) which appear in the main decomposition (8.1).

Lemma B.3. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{W}_n = 2\sigma^2 \quad \text{a.s.} \quad (\text{B.3})$$

Proof : First of all, we have the decomposition $\mathcal{W}_{n+1} = \mathcal{T}_{n+1} + \mathcal{R}_{n+1}$ where

$$\begin{aligned} \mathcal{T}_{n+1} &= \sum_{k=p}^n \frac{\Delta M_{k+1}^t \Lambda^{-1} \Delta M_{k+1}}{|\mathbb{T}_k|}, \\ \mathcal{R}_{n+1} &= \sum_{k=p}^n \frac{\Delta M_{k+1}^t (|\mathbb{T}_k| \Sigma_k^{-1} - \Lambda^{-1}) \Delta M_{k+1}}{|\mathbb{T}_k|}. \end{aligned}$$

We claim that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{F}_n = (p+1)\sigma^2 \quad \text{a.s.}$$

It will ensure via (8.5) that $\mathcal{R}_n = o(n)$ a.s. leading to (B.3). One can observe that $\mathcal{T}_{n+1} = \text{tr}(\Lambda^{-1/2} H_{n+1} \Lambda^{-1/2})$ where

$$H_{n+1} = \sum_{k=p}^n \frac{\Delta M_{k+1} \Delta M_{k+1}^t}{|\mathbb{T}_k|}.$$

Our goal is to make use of the strong law of large numbers for martingale transforms, so we start by adding and subtracting a term involving the conditional expectation of ΔH_{n+1} given \mathcal{F}_n . We have already seen in Section 4 that for all $n \geq p-1$, $\mathbb{E}[\Delta M_{n+1} \Delta M_{n+1}^t | \mathcal{F}_n] = \Gamma \otimes \Phi_n \Phi_n^t$. Consequently, we can split H_{n+1} into two terms

$$H_{n+1} = \sum_{k=p}^n \frac{\Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|} + K_{n+1}$$

where

$$K_{n+1} = \sum_{k=p}^n \frac{\Delta M_{k+1} \Delta M_{k+1}^t - \Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|}.$$

On the one hand, it follows from convergence (5.1) and Lemma A.2 that

$$\lim_{n \rightarrow +\infty} \frac{\Phi_n \Phi_n^t}{|\mathbb{T}_n|} = \frac{1}{2} L \quad \text{a.s.}$$

Thus, Cesaro convergence yields

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{k=p}^n \frac{\Gamma \otimes \Phi_k \Phi_k^t}{|\mathbb{T}_k|} = \frac{1}{2} (\Gamma \otimes L) \quad \text{a.s.} \quad (\text{B.4})$$

On the other hand, the sequence (K_n) is obviously a matrix martingale transform satisfying

$$\Delta K_{n+1} = K_{n+1} - K_n = \frac{1}{|\mathbb{T}_{n+1}|} \sum_{i,j \in \mathbb{G}_n} \Gamma_{ij} \otimes \begin{pmatrix} 1 & \mathbb{X}_j^t \\ \mathbb{X}_i & \mathbb{X}_i \mathbb{X}_j^t \end{pmatrix}$$

where

$$\Gamma_{ij} = \begin{pmatrix} \varepsilon_{2i} \varepsilon_{2j} - \mathbb{I}_{i=j} \sigma^2 & \varepsilon_{2i} \varepsilon_{2j+1} - \mathbb{I}_{i=j} \rho \\ \varepsilon_{2i+1} \varepsilon_{2j} - \mathbb{I}_{i=j} \rho & \varepsilon_{2i+1} \varepsilon_{2j+1} - \mathbb{I}_{i=j} \sigma^2 \end{pmatrix}.$$

For all $u \in \mathbb{R}^{2(p+1)}$, let $K_n(u) = u^t K_n u$. It follows from tedious but straightforward calculations, together with (A.4), (A.13) and the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [4] that $K_n(u) = o(n)$ a.s. for all $u \in \mathbb{R}^{2(p+1)}$ leading to $K_n = o(n)$ a.s. Hence, we infer from (B.4) that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} H_n = \frac{1}{2} (\Gamma \otimes L) \quad \text{a.s.} \quad (\text{B.5})$$

Finally, we find from (B.5) that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{G}_n &= \frac{1}{2} \text{tr}(\Lambda^{-1/2}(\Gamma \otimes L)\Lambda^{-1/2}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}((\Gamma \otimes L)\Lambda^{-1}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}(\Gamma \otimes I_{p+1}) = (p+1)\sigma^2 \quad \text{a.s.} \end{aligned}$$

which completes the proof of Lemma B.3 □

Lemma B.4. *Assume that (ε_n) satisfies (H.1) to (H.3). Then, we have*

$$\mathcal{B}_{n+1} = o(n) \quad \text{a.s.}$$

Proof : Recall that

$$\mathcal{B}_{n+1} = 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Delta M_{k+1} = 2 \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Psi_k \xi_{k+1}.$$

The sequence (\mathcal{B}_n) is a real martingale transform satisfying

$$\Delta \mathcal{B}_{n+1} = \mathcal{B}_{n+1} - \mathcal{B}_n = 2M_n^t \Sigma_n^{-1} \Psi_n \xi_{n+1}.$$

Consequently, via the strong law of large numbers for martingale transforms [4], we find that either (\mathcal{B}_n) converges a.s. or $\mathcal{B}_{n+1} = o(v_n)$ a.s. where

$$v_n = \sum_{k=p}^n M_k^t \Sigma_k^{-1} \Psi_k \Psi_k^t \Sigma_k^{-1} M_k.$$

However, for all $n \geq 2^{p-1}$, $\Psi_n \Psi_n^t = I_2 \otimes \Phi_n \Phi_n^t$ which implies that

$$v_n = \sum_{k=p}^n M_k^t \Sigma_k^{-1} (I_2 \otimes \Phi_k \Phi_k^t) \Sigma_k^{-1} M_k = \sum_{k=p}^n M_k^t (I_2 \otimes S_k^{-1} \Phi_k \Phi_k^t S_k^{-1}) M_k.$$

Furthermore, it follows from Lemma B.1 that

$$S_{n-1}^{-1} - S_n^{-1} = S_n^{-1} \Phi_n (I_{\delta_n} + l_n) \Phi_n^t S_n^{-1} \geq S_n^{-1} \Phi_n \Phi_n^t S_n^{-1}$$

as the matrix l_n is definite positive. Therefore, we obtain that

$$v_n \leq \sum_{k=p}^n M_k^t (\Sigma_{k-1}^{-1} - \Sigma_k^{-1}) M_k = \mathcal{A}_n.$$

Finally, we deduce from the main decomposition (8.1) that

$$\mathcal{V}_{n+1} + \mathcal{A}_n = o(\mathcal{A}_n) + \mathcal{O}(n) \quad \text{a.s.}$$

leading to $\mathcal{V}_{n+1} = \mathcal{O}(n)$ and $\mathcal{A}_n = \mathcal{O}(n)$ a.s. as \mathcal{V}_{n+1} and \mathcal{A}_n are non-negative, which implies in turn that $\mathcal{B}_n = o(n)$ a.s. completing the proof of Lemma B.4. □

Proof of Lemma 8.1 : Convergence (8.2) immediately follows from (8.1) together with Lemmas B.3 and B.4. □

C On Wei's Lemma

In order to prove (8.3), we shall apply Wei's Lemma given in [13] page 1672, to each entry of the vector-valued martingale

$$M_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \begin{pmatrix} \varepsilon_{2i} \\ \mathbb{X}_i \varepsilon_{2i} \\ \varepsilon_{2i+1} \\ \mathbb{X}_i \varepsilon_{2i+1} \end{pmatrix}.$$

We shall only carry out the proof for the first $(p+1)$ of M_n inasmuch as the proof for the $(p+1)$ last components follows exactly the same lines. Denote

$$P_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \varepsilon_{2i} \quad \text{and} \quad Q_n = \sum_{k=p}^n \sum_{i \in \mathbb{G}_{k-1}} \mathbb{X}_i \varepsilon_{2i}.$$

On the one hand, P_n can be rewritten as $P_n = \sum_{k=p}^n \sqrt{|\mathbb{G}_{k-1}|} v_k$ where

$$v_n = \frac{1}{\sqrt{|\mathbb{G}_{n-1}|}} \sum_{i \in \mathbb{G}_{n-1}} \varepsilon_{2i}.$$

We clearly have $\mathbb{E}[v_{n+1} | \mathcal{F}_n] = 0$, $\mathbb{E}[v_{n+1}^2 | \mathcal{F}_n] = \sigma^2$ a.s. Moreover, it follows from (H.1) to (H.3) together with Cauchy-Schwarz inequality that

$$\begin{aligned} \mathbb{E}[v_{n+1}^4 | \mathcal{F}_n] &= \frac{1}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] + \frac{3}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \sum_{j \neq i} \mathbb{E}[\varepsilon_{2i}^2 | \mathcal{F}_n] \mathbb{E}[\varepsilon_{2j}^2 | \mathcal{F}_n] \\ &\leq 3 \sup_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] \quad \text{a.s.} \end{aligned}$$

which implies that $\sup \mathbb{E}[v_{n+1}^4 | \mathcal{F}_n] < +\infty$ a.s. Consequently, we deduce from Wei's Lemma that for all $\delta > 1/2$,

$$P_n^2 = o(|\mathbb{T}_{n-1}| n^\delta) \quad \text{a.s.}$$

On the other hand, we also have $Q_n = \sum_{k=p}^n \sqrt{|\mathbb{G}_{k-1}|} w_k$ where

$$w_n = \frac{1}{\sqrt{|\mathbb{G}_{n-1}|}} \sum_{i \in \mathbb{G}_{n-1}} \mathbb{X}_i \varepsilon_{2i}.$$

It is not hard to see that $\mathbb{E}[w_{n+1} | \mathcal{F}_n] = 0$ a.s. Moreover, for all $1 \leq k \leq p$, let $w_n(k)$ be the k th coordinate of the vector w_n . It follows from (H.1) to (H.3) and Cauchy-Schwarz inequality that for all $1 \leq k \leq p$,

$$\begin{aligned} \mathbb{E}[w_{n+1}(k)^4 | \mathcal{F}_n] &\leq \frac{1}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^4 \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] + \frac{3\sigma^4}{|\mathbb{G}_n|^2} \sum_{i \in \mathbb{G}_n} \sum_{j \neq i} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^2 X_{\lfloor \frac{j}{2^{k-1}} \rfloor}^2 \\ &\leq 3 \sup_{i \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2i}^4 | \mathcal{F}_n] \left(\frac{1}{|\mathbb{G}_n|} \sum_{i \in \mathbb{G}_n} X_{\lfloor \frac{i}{2^{k-1}} \rfloor}^2 \right)^2 \quad \text{a.s.} \end{aligned}$$

Hence, we obtain from Lemma 7.2 that for all $1 \leq k \leq p$, $\sup \mathbb{E}[w_{n+1}(k)^4 | \mathcal{F}_n] < +\infty$ a.s. Once again, we deduce from Wei's Lemma applied to each component of Q_n that for all $\delta > 1/2$,

$$\|Q_n\|^2 = o(|\mathbb{T}_{n-1}|n^\delta) \quad \text{a.s.}$$

which completes the proof of (8.3). □

D On the convergence of the covariance estimator

It remains to prove that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k})(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) = \lim_{n \rightarrow \infty} \frac{R_n}{2n} = (p+1)\rho \quad \text{a.s.}$$

where

$$R_n = \sum_{k \in \mathbb{T}_{n-1, p-1}} (\widehat{V}_k - V_k)^t J_2 (\widehat{V}_k - V_k).$$

It is not possible to make use of the previous convergence (9.1) because the matrix

$$J_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

is not positive definite. Hence, it is necessary to rewrite our proofs. Denote

$$\mathcal{V}'_n = M_n^t \Sigma_{n-1}^{-1/2} (J_2 \otimes I_{p+1}) \Sigma_{n-1}^{-1/2} M_n.$$

As in the proof of Theorem 5.2, we have the decomposition

$$\mathcal{V}'_{n+1} + \mathcal{A}'_n = \mathcal{V}'_1 + \mathcal{B}'_{n+1} + \mathcal{W}'_{n+1} \quad (\text{D.1})$$

where

$$\begin{aligned} \mathcal{A}'_n &= \sum_{k=p}^n M_k^t (J_2 \otimes (S_{k-1}^{-1} - S_k^{-1})) M_k, \\ \mathcal{B}'_{n+1} &= 2 \sum_{k=p}^n M_k^t (J_2 \otimes S_k^{-1}) \Delta M_{k+1}, \\ \mathcal{W}'_{n+1} &= \sum_{k=p}^n \Delta M_{k+1}^t (J_2 \otimes S_k^{-1}) \Delta M_{k+1}. \end{aligned}$$

First of all, via the same lines as in Appendix B, we obtain that

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{W}'_n &= \frac{1}{2} \text{tr}((J_2 \otimes L^{-1})^{1/2} (\Gamma \otimes L) (J_2 \otimes L^{-1})^{1/2}) \quad \text{a.s.} \\ &= \frac{1}{2} \text{tr}(\Gamma J_2 \otimes I_{p+1}) = (p+1)\rho \quad \text{a.s.} \end{aligned}$$

Next, (\mathcal{B}'_n) is a real martingale transform satisfying $\mathcal{B}'_{n+1} = o(n)$ a.s. Hence, we find the analogous of convergence (8.2)

$$\lim_{n \rightarrow +\infty} \frac{\mathcal{V}'_{n+1} + \mathcal{A}'_n}{n} = (p+1)\rho \quad \text{a.s.} \quad (\text{D.2})$$

Furthermore, it follows from Wei's Lemma that for all $\delta > 1/2$,

$$\mathcal{V}'_n = o(n^\delta) \quad \text{a.s.} \quad (\text{D.3})$$

Therefore, we infer (D.1), (D.2) and (D.3) that

$$\lim_{n \rightarrow +\infty} \frac{1}{n} \mathcal{A}'_n = (p+1)\rho \quad \text{a.s.} \quad (\text{D.4})$$

Finally, by the same lines as in the proof of the first part of Theorem 5.3, we find that

$$\lim_{n \rightarrow \infty} \frac{R_n}{n} = 2 \lim_{n \rightarrow \infty} \frac{\mathcal{A}'_n}{n} = 2(p+1)\rho \quad \text{a.s.}$$

which completes the proof of convergence (9.2). \square

Acknowledgements

The authors would like to thanks the anonymous referees for their very careful reading of the manuscript and for their suggestion to extend the first version of the paper to asymmetric BAR(p) processes.

References

- [1] BASAWA, I. V., AND ZHOU, J. Non-Gaussian bifurcating models and quasi-likelihood estimation. *J. Appl. Probab.* 41A (2004), 55–64. MR2057565
- [2] COWAN, R., AND STAUDTE, R. G. The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42 (1986), 769–783.
- [3] DELMAS, J.-F., AND MARSALLE, L. Detection of cellular aging in a Galton-Watson process. *arXiv*, 0807.0749 (2008).
- [4] DUFLO, M. *Random iterative models*, vol. 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997. MR1485774
- [5] GUYON, J. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* 17, 5-6 (2007), 1538–1569. MR2358633
- [6] GUYON, J., BIZE, A., PAUL, G., STEWART, E., DELMAS, J.-F., AND TADDÉI, F. Statistical study of cellular aging. In *CEMRACS 2004—mathematics and applications to biology and medicine*, vol. 14 of *ESAIM Proc.* EDP Sci., Les Ulis, 2005, pp. 100–114 (electronic). MR2226805
- [7] HALL, P., AND HEYDE, C. C. *Martingale limit theory and its application*. Academic Press Inc., New York, 1980. Probability and Mathematical Statistics. MR0624435

- [8] HAMILTON, J. D. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994. MR1278033
- [9] HUGGINS, R. M. Robust inference for variance components models for single trees of cell lineage data. *Ann. Statist.* 24, 3 (1996), 1145–1160. MR1401842
- [10] HUGGINS, R. M., AND BASAWA, I. V. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* 36, 4 (1999), 1225–1233. MR1746406
- [11] HUGGINS, R. M., AND BASAWA, I. V. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* 42, 4 (2000), 423–432. MR1802966
- [12] HWANG, S. Y., BASAWA, I. V., AND YEO, I. K. Local asymptotic normality for bifurcating autoregressive processes and related asymptotic inference. *Statistical Methodology* 6 (2009), 61–69.
- [13] WEI, C. Z. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.* 15, 4 (1987), 1667–1682. MR0913581
- [14] ZHOU, J., AND BASAWA, I. V. Least-squares estimation for bifurcating autoregressive processes. *Statist. Probab. Lett.* 74, 1 (2005), 77–88. MR2189078
- [15] ZHOU, J., AND BASAWA, I. V. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Ser. Anal.* 26, 6 (2005), 825–842. MR2203513

- 1.2 [dSGP⁺11] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Parameters estimation for bifurcating autoregressive processes with missing data. *Electronic Journal of Statistics*, 5 :1313–1353, 2011.

Parameters estimation for asymmetric bifurcating autoregressive processes with missing data

Benoîte de Saporta

Université de Bordeaux, GREThA CNRS UMR 5113, IMB CNRS UMR 5251

and INRIA Bordeaux Sud Ouest team CQFD, France

e-mail: saporta@math.u-bordeaux1.fr

Anne Gégout-Petit

Université de Bordeaux, IMB, CNRS UMR 525

and INRIA Bordeaux Sud Ouest team CQFD, France

e-mail: anne.petit@u-bordeaux2.fr

and

Laurence Marsalle

Université de Lille 1, Laboratoire Paul Painlevé, CNRS UMR 8524, France

e-mail: laurence.marsalle@univ-lille1.fr

Abstract: We estimate the unknown parameters of an asymmetric bifurcating autoregressive process (BAR) when some of the data are missing. In this aim, we model the observed data by a two-type Galton-Watson process consistent with the binary tree structure of the data. Under independence between the process leading to the missing data and the BAR process and suitable assumptions on the driven noise, we establish the strong consistency of our estimators on the set of non-extinction of the Galton-Watson process, via a martingale approach. We also prove a quadratic strong law and the asymptotic normality.

AMS 2000 subject classifications: Primary 62F12, 62M09; secondary 60J80, 92D25, 60G42.

Keywords and phrases: Least squares estimation, bifurcating autoregressive process, missing data, Galton-Watson process, joint model, martingales, limit theorems.

Received June 2011.

1. Introduction

Bifurcating autoregressive processes (BAR) generalize autoregressive (AR) processes, when the data have a binary tree structure. Typically, they are involved in modeling cell lineage data, since each cell in one generation gives birth to two offspring in the next one. Cell lineage data usually consist of observations of some quantitative characteristic of the cells, over several generations descended

from an initial cell. BAR processes take into account both inherited and environmental effects to explain the evolution of the quantitative characteristic under study. They were first introduced by Cowan and Staudte [4]. In their paper, the original BAR process was defined as follows. The initial cell is labelled 1, and the two offspring of cell k are labelled $2k$ and $2k + 1$. If X_k denotes the quantitative characteristic of individual k , then the first-order BAR process is given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} &= a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} &= a + bX_k + \varepsilon_{2k+1}. \end{cases}$$

The noise sequence $(\varepsilon_{2k}, \varepsilon_{2k+1})$ represents environmental effects, while a, b are unknown real parameters, with $|b| < 1$, related to the inherited effects. The driven noise $(\varepsilon_{2k}, \varepsilon_{2k+1})$ was originally supposed to be independent and identically distributed with normal distribution. But since two sister cells are in the same environment at their birth, ε_{2k} and ε_{2k+1} are allowed to be correlated, inducing a correlation between sister cells, distinct from the correlation inherited from their mother.

Recently, experiments made by biologists on aging of *Escherichia coli* [15], motivated mathematical and statistical studies of the asymmetric BAR process, that is when the quantitative characteristics of the even and odd sisters are allowed to depend on their mother's through different sets of parameters (a, b) , see Equation (2.1) below. In [9, 8], Guyon proposes an interpretation of the asymmetric BAR process as a bifurcating Markov chain, which allows him to derive laws of large numbers and central limit theorems for the least squares estimators of the unknown parameters of the process. This Markov chain approach was further developed by Bansaye [2] in the context of cell division with parasite infection, and by Delmas and Marsalle [5], where the cells are allowed to die. Another approach based on martingales theory was proposed by Bercu, de Saporta and Gégout-Petit [3], to sharpen the asymptotic analysis of Guyon under weaker assumptions.

The originality of this paper is that we take into account possibly missing data in the estimation procedure of the parameters of the asymmetric BAR process, see Figure 1 for an example. This is a problem of practical interest, as experimental data are often incomplete, either because some cells died, or because the measurement of the characteristic under study was impossible or faulty. For instance, among the 94 colonies dividing up to 9 times studied in [15], in average, there are about 47% of missing data. It is important to take this phenomenon into account in the model for a rigorous statistical study.

Missing data in bifurcating processes were first modeled by Delmas and Marsalle [5]. They defined the genealogy of the cells through a Galton-Watson process, but they took into account the possible asymmetry problem only by differentiating the reproduction laws according to the daughter's type (even or odd). The bifurcating process was thus still a Markov chain. However, considering the biological issue of aging in *E. coli* naturally leads to introduce the possibility that two cells of different types may not have the same reproduction law. In this paper, we thus introduce a two-type Galton-Watson process to

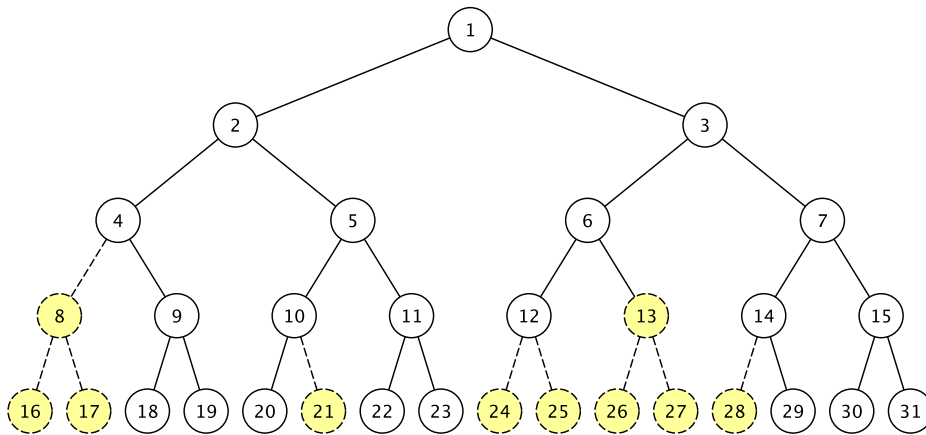


FIG 1. A tree associated with the bifurcating auto-regressive process up to the 4th generation. The dashed cells are not observed.

model the genealogy, and lose the Markovian structure of the bifurcating chain, so that we cannot use the same approach as [5]. Instead, we use the martingale approach introduced in [3]. It must be pointed out that missing data are not dealt with in [3], so that we cannot directly use their results either. In particular, the observation process is another source of randomness that requires stronger moment assumptions on the driven noise of the BAR process and careful choice between various filtrations. In addition, the normalizing terms are now random and the convergences are only available on the random non-extinction set of the observed process.

The *naive* approach to handle missing data would be to replace the sums over all data in the estimators by sums over the observed data only. Our approach is slightly more subtle, as we distinguish whether a cell has even or odd daughters. We propose a joint model where the structure for the observed data is based on a two-type Galton-Watson process consistent with the possibly asymmetric structure of the BAR process. See e.g. [12, 1, 10] for a presentation of multi-type Galton-Watson processes and general branching processes. Note also that our estimation procedure does not require the previous knowledge of the parameters of the two-type Galton-Watson process.

This paper is organized as follows. In Section 2, we first introduce our BAR model as well as related notation, then we define and recall results on the two-type Galton-Watson process used to model the observation process. In Section 3, we give the least square estimator for the parameters of observed BAR process and we state our main results on the convergence and asymptotic normality of our estimators as well as estimation results on data. The proofs are detailed in the following sections.

2. Joint model

We now introduce our joint model, starting with the asymmetric BAR process for the variables of interest.

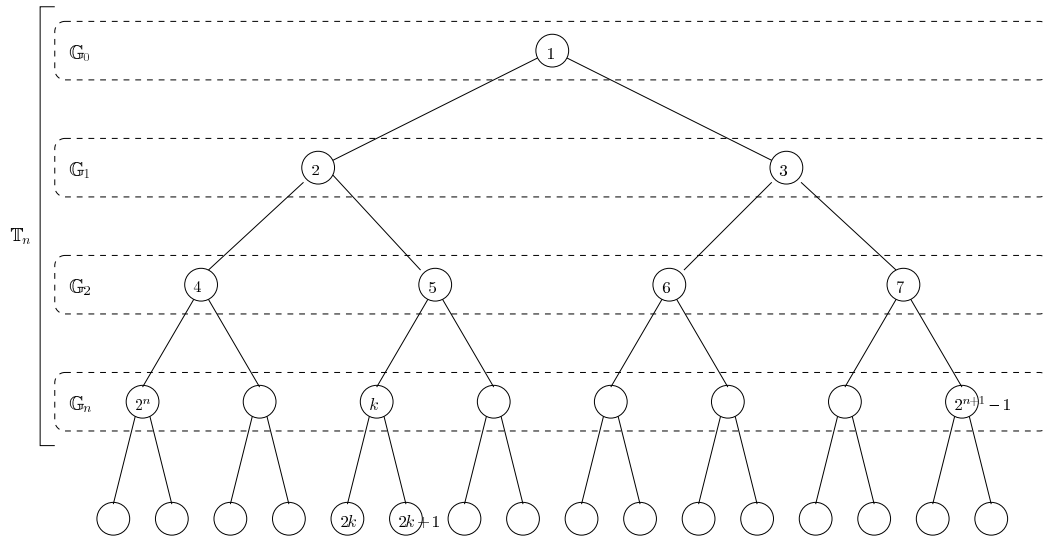


FIG 2. The tree associated with the bifurcating auto-regressive process.

2.1. Bifurcating autoregressive processes

On the probability space $(\Omega, \mathcal{A}, \mathbb{P})$, we consider the first-order asymmetric BAR process given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} &= a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} &= c + dX_k + \varepsilon_{2k+1}. \end{cases} \tag{2.1}$$

The initial state X_1 is the characteristic of the ancestor, while $(\varepsilon_{2k}, \varepsilon_{2k+1})$ is the driven noise of the process. In all the sequel, we shall assume that $\mathbb{E}[X_1^8] < \infty$. Moreover, as in the previous literature, the parameters (a, b, c, d) belong to \mathbb{R}^4 with

$$0 < \max(|b|, |d|) < 1.$$

This assumption ensures the stability (non explosion) of the BAR process. As explained in the introduction, one can see this BAR process as a first-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor, see Figure 2 for an illustration. We use the same notation as in [3]. For all $n \geq 1$, denote the n -th generation by $\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}$. In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation, and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the first ancestor. Let \mathbb{G}_{r_k} be the generation of individual k , which means that $r_k = \lceil \log_2(k) \rceil$, where $\lceil x \rceil$ denotes the largest integer less than or equal to x . Recall that the two offspring of individual k are labelled $2k$ and $2k + 1$, or conversely, the mother of individual k is $\lfloor k/2 \rfloor$. More generally, the ancestors of individual k are $\lfloor k/2 \rfloor, \lfloor k/2^2 \rfloor, \dots, \lfloor k/2^{r_k} \rfloor$. Denote by $\mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$, the sub-tree of all individuals from the original individual up to the n -th generation. Note that the cardinality $|\mathbb{G}_n|$ of \mathbb{G}_n is 2^n , while that of \mathbb{T}_n is $|\mathbb{T}_n| = 2^{n+1} - 1$. Next, \mathbb{T} denotes the complete tree, so to speak $\mathbb{T} = \bigcup_{n \geq 0} \mathbb{G}_n = \bigcup_{n \geq 0} \mathbb{T}_n = \mathbb{N}^* = \mathbb{N} \setminus \{0\}$.

Finally, we need to distinguish the individuals in \mathbb{G}_n and \mathbb{T}_n according to their type. Since we are dealing with the types even and odd, that we will also label 0 and 1, we set

$$\begin{aligned} \mathbb{G}_n^0 &= \mathbb{G}_n \cap (2\mathbb{N}), & \mathbb{G}_n^1 &= \mathbb{G}_n \cap (2\mathbb{N} + 1), & \mathbb{T}_n^0 &= \mathbb{T}_n \cap (2\mathbb{N}), \\ \mathbb{T}_n^1 &= \mathbb{T}_n \cap (2\mathbb{N} + 1), & \mathbb{T}^0 &= \mathbb{T} \cap (2\mathbb{N}) & \text{and} & \mathbb{T}^1 = \mathbb{T} \cap (2\mathbb{N} + 1). \end{aligned} \quad (2.2)$$

We now state our assumptions on the noise sequence. Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the first-order BAR process, which means that \mathcal{F}_n is the σ -algebra generated by all individuals up to the n -th generation, $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$. In all the sequel, we shall make use of the following moment and independence hypotheses.

(HN.1) For all $n \geq 0$ and for all $k \in \mathbb{G}_{n+1}$, ε_k belongs to L^8 . Moreover, there exist $(\sigma^2, \tau^4, \kappa^8) \in (0, +\infty)^3$, $(|\rho'|, \nu^2, \lambda^4) \in [0, 1]^3$ such that:

- $\forall n \geq 0$ and $k \in \mathbb{G}_{n+1}$,

$$\mathbb{E}[\varepsilon_k | \mathcal{F}_n] = 0, \quad \mathbb{E}[\varepsilon_k^2 | \mathcal{F}_n] = \sigma^2, \quad \mathbb{E}[\varepsilon_k^4 | \mathcal{F}_n] = \tau^4, \quad \mathbb{E}[\varepsilon_k^8 | \mathcal{F}_n] = \kappa^8 \quad \text{a.s.}$$

- $\forall n \geq 0 \quad \forall k \neq l \in \mathbb{G}_{n+1}$ with $[k/2] = [l/2]$,

$$\mathbb{E}[\varepsilon_k \varepsilon_l | \mathcal{F}_n] = \rho = \rho' \sigma^2, \quad \mathbb{E}[\varepsilon_{2k}^2 \varepsilon_{2k+1}^2 | \mathcal{F}_n] = \nu^2 \tau^4, \quad \mathbb{E}[\varepsilon_{2k}^4 \varepsilon_{2k+1}^4 | \mathcal{F}_n] = \lambda^4 \kappa^8 \quad \text{a.s.}$$

(HN.2) For all $n \geq 0$ the random vectors $\{(\varepsilon_{2k}, \varepsilon_{2k+1}), k \in \mathbb{G}_n\}$ are conditionally independent given \mathcal{F}_n .

2.2. Observation process

We now turn to the modeling of the observation process. The observation process is intended to encode if a datum is missing or not. The natural property it has thus to satisfy is the following: if the datum is missing for some individual, it is also missing for all its descendants. Indeed, the datum may be missing because of the death of the individual, or because the individual is the last of its lineage at the end of the data's gathering, see Figure 3 for an example of partially observed tree.

2.2.1. Definition of the observation process

Mathematically, we define the observation process, $(\delta_k)_{k \in \mathbb{T}}$, as follows. We set $\delta_1 = 1$ and define recursively the sequence through the following equalities:

$$\delta_{2k} = \delta_k \zeta_k^0 \quad \text{and} \quad \delta_{2k+1} = \delta_k \zeta_k^1, \quad (2.3)$$

where $(\zeta_k = (\zeta_k^0, \zeta_k^1))$ is a sequence of independent random vectors of $\{0, 1\}^2$, ζ_k^i standing for the number (0 or 1) of descendants of type i of individual k . The sequences $(\zeta_k, k \in 2\mathbb{N}^*)$ and $(\zeta_k, k \in 2\mathbb{N} + 1)$ are sequences of identically

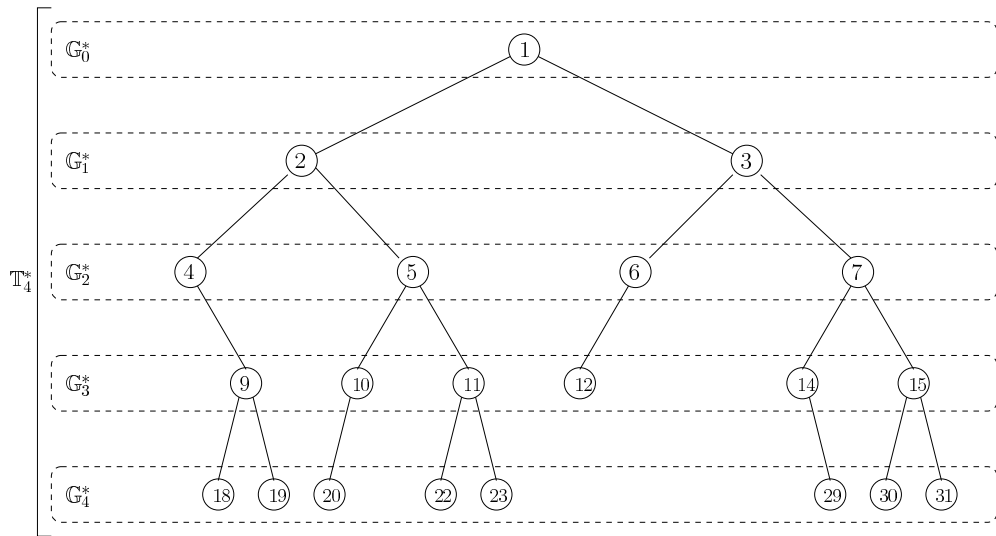


FIG 3. The tree associated with the observed data of the tree in Figure 1.

distributed random vectors. We specify the common laws of these two sequences using their generating functions, $f^{(0)}$ and $f^{(1)}$ respectively:

$$\begin{aligned} f^{(0)}(s_0, s_1) &= p^{(0)}(0, 0) + p^{(0)}(1, 0)s_0 + p^{(0)}(0, 1)s_1 + p^{(0)}(1, 1)s_0s_1, \\ f^{(1)}(s_0, s_1) &= p^{(1)}(0, 0) + p^{(1)}(1, 0)s_0 + p^{(1)}(0, 1)s_1 + p^{(1)}(1, 1)s_0s_1, \end{aligned}$$

where $p^{(i)}(j_0, j_1)$ is the probability that an individual of type i gives birth to j_0 descendants of type 0, and j_1 of type 1. The sequence (δ_k) is thus completely defined. We also assume that the observation process is independent from the BAR process.

(HI) The sequences (δ_k) and (ζ_k) are independent from the sequences (X_k) and (ε_k) .

Remark that, since both ζ_k^0 and ζ_k^1 take values in $\{0, 1\}$ for all k , the observation process (δ_k) is itself taking values in $\{0, 1\}$. Finally, Equation (2.3) ensures that if $\delta_k = 0$ for some $k \geq 2$, then for all its descendants j , $\delta_j = 0$. In relation with the observation process (δ_k) , we introduce two filtrations: $\mathcal{Z}_n = \sigma\{\zeta_k, k \in \mathbb{T}_n\}$, $\mathcal{O}_n = \sigma\{\delta_k, k \in \mathbb{T}_n\}$, and the sigma field $\mathcal{O} = \sigma\{\delta_k, k \in \mathbb{T}\}$. Notice that $\mathcal{O}_{n+1} \subset \mathcal{Z}_n$. We also define the sets of observed individuals as follows:

$$\mathbb{G}_n^* = \{k \in \mathbb{G}_n : \delta_k = 1\} \quad \text{and} \quad \mathbb{T}_n^* = \{k \in \mathbb{T}_n : \delta_k = 1\}.$$

Finally, let \mathcal{E} be the event corresponding to the cases when there are no individual left to observe. More precisely,

$$\mathcal{E} = \bigcup_{n \geq 1} \{|\mathbb{G}_n^*| = 0\}. \tag{2.4}$$

We will denote $\overline{\mathcal{E}}$ the complementary set of \mathcal{E} .

2.2.2. Results on the observation process

Let us introduce some additional notation. For $n \geq 1$, we define the number of observed individuals among the n -th generation, distinguishing according to their types:

$$Z_n^0 = |\mathbb{G}_n^* \cap 2\mathbb{N}| \quad \text{and} \quad Z_n^1 = |\mathbb{G}_n^* \cap (2\mathbb{N} + 1)|, \tag{2.5}$$

and we set, for all $n \geq 1$, $\mathbf{Z}_n = (Z_n^0, Z_n^1)$. Note that for $i \in \{0, 1\}$ and $n \geq 1$ one has

$$Z_n^i = \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i}.$$

One has $\mathbb{G}_0^* = \mathbb{G}_0 = \{1\}$, but, even if 1 is odd, the individual whose lineage we study may as well be of type 0 as of type 1. Consequently, we will work with possibly two different initial laws: $\mathbb{P}(\cdot | \mathbf{Z}_0 = \mathbf{e}_i)$, for $i \in \{0, 1\}$, where $\mathbf{e}_0 = (1, 0)$ and $\mathbf{e}_1 = (0, 1)$. The process $(\mathbf{Z}_n, n \geq 0)$ is thus a two-type Galton-Watson process, and all the results we are giving in this section mainly come from [12]. Notice that the law of ζ_k , for even k , is the law of reproduction of an individual of type 0, the first component of ζ_k giving the number of children of type 0, the second the number of children of type 1. The same holds for ζ_k , with odd k , *mutatis mutandis*. This ensures the existence of moments of all order for these reproduction laws, and we can thus define the descendants matrix \mathbf{P}

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

where $p_{i0} = p^{(i)}(1, 0) + p^{(i)}(1, 1)$ and $p_{i1} = p^{(i)}(0, 1) + p^{(i)}(1, 1)$, for $i \in \{0, 1\}$. The quantity $p_{ij} = \mathbb{E}[\zeta_{2+i}^j]$ is thus the expected number of descendants of type j of an individual of type i . We also introduce the variance of the laws of reproduction: $\sigma_{ij}^2 = \mathbb{E}[(\zeta_{2+i}^j - p_{ij})^2]$, for $(i, j) \in \{0, 1\}^2$. Note that $\sigma_{ij}^2 = p_{ij}(1 - p_{ij})$. It is well-known (see e.g. Theorem 5.1 of [12]) that when all the entries of the matrix \mathbf{P} are positive, \mathbf{P} has a positive strictly dominant eigenvalue, denoted π , which is also simple. We make the following main assumptions on the matrix \mathbf{P} .

(HO) All entries of the matrix \mathbf{P} are positive: for all $(i, j) \in \{0, 1\}^2$, $p_{ij} > 0$, and the dominant eigenvalue is greater than one: $\pi > 1$.

Hence, still following Theorem 5.1 of [12], we know that there exist left and right eigenvectors for π which are positive, in the sense that each component of the vector is positive. We call $\mathbf{y} = (y^0, y^1)^t$ such a right eigenvector, and $\mathbf{z} = (z^0, z^1)$ such a left one; without loss of generality, we choose \mathbf{z} such that $z^0 + z^1 = 1$. Regarding the two-type Galton-Watson process (\mathbf{Z}_n) , π plays the same role as the expected number of offspring, in the case of standard Galton-Watson processes. In particular, π is related to the extinction of the process, where the set of extinction of (\mathbf{Z}_n) is defined as $\cup_{n \geq 1} \{\mathbf{Z}_n = (0, 0)\}$. Notice that $\{\mathbf{Z}_n = (0, 0)\} = \{Z_n^0 + Z_n^1 = 0\} = \{|\mathbb{G}_n^*| = 0\}$, so that this set coincides with \mathcal{E} , defined by Eq. (2.4). Now let $\mathbf{q} = (q^0, q^1)$, where, for $i \in \{0, 1\}$,

$$q^i = \mathbb{P}(\mathcal{E} | \mathbf{Z}_0 = \mathbf{e}_i).$$

The probability q^i is thus the extinction probability if initially there is one individual of type i . These two probabilities allow to compute the extinction probability under any initial distribution, since $\mathbb{P}(\mathcal{E}) = \mathbb{E}[(q^0)^{Z_0^0} (q^1)^{Z_0^1}]$, thanks to the branching property. Hypothesis **(HO)** means that the Galton-Watson process (Z_n) is super-critical, and ensures that $0 \leq q^i < 1$, for both $i = 0$ and $i = 1$. This immediately yields

$$\mathbb{P}(\mathcal{E}) < 1. \tag{2.6}$$

Under that condition, we also have the existence of a non-negative random variable W such that for any initial distribution of Z_0

$$\lim_{n \rightarrow +\infty} \frac{Z_n}{\pi^n} = \lim_{n \rightarrow +\infty} \frac{\pi - 1}{\pi^{n+1} - 1} \sum_{\ell=0}^n Z_\ell = Wz \quad \text{a.s.} \tag{2.7}$$

It is well-known that $\{W = 0\} = \mathcal{E}$ a.s., so that the set $\{W > 0\}$ can be viewed as the set of non-extinction $\bar{\mathcal{E}}$ of (Z_n) , up to a negligible set. These results give the asymptotic behavior of the number of observed individuals, since $|\mathbb{G}_n^*| = Z_n^0 + Z_n^1$, and $|\mathbb{T}_n^*| = \sum_{\ell=0}^n (Z_\ell^0 + Z_\ell^1)$:

$$\lim_{n \rightarrow +\infty} \frac{|\mathbb{G}_n^*|}{\pi^n} = \lim_{n \rightarrow +\infty} \frac{\pi - 1}{\pi^{n+1} - 1} |\mathbb{T}_n^*| = W \quad \text{a.s.}$$

Roughly speaking, this means that π^n is a deterministic equivalent of $|\mathbb{T}_n^*|$ and Eq. (2.7) implies that z^i is the asymptotic proportion of cells of type i in a given generation. We will thus very often replace $|\mathbb{T}_n^*|$ by π^n for computations, and the next lemma will be used frequently to replace π^n by $|\mathbb{T}_n^*|$.

Lemma 2.1. *Under assumption (HO), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\pi^n}{|\mathbb{T}_n^*|} = \frac{\pi - 1}{\pi} \frac{1}{W} \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

2.3. Joint model

The model under study in this paper is therefore the observed BAR process defined by

$$\begin{cases} \delta_{2k} X_{2k} &= \delta_{2k} \quad (a + bX_k + \varepsilon_{2k}), \\ \delta_{2k+1} X_{2k+1} &= \delta_{2k+1} \quad (c + dX_k + \varepsilon_{2k+1}). \end{cases}$$

The aim of this paper is to study the sharp asymptotic properties of the least-squares estimators of the parameters (a, b, c, d) and the variance matrix of the noise process.

3. Least-squares estimation

Our goal is to estimate $\theta = (a, b, c, d)^t$ from the observed individuals up to the n -th generation, that is the observed sub-tree \mathbb{T}_n^* .

3.1. Definition of the estimators

We propose to make use of the standard least-squares (LS) estimator $\widehat{\boldsymbol{\theta}}_n$ which minimizes

$$\Delta_n(\boldsymbol{\theta}) = \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k}(X_{2k} - a - bX_k)^2 + \delta_{2k+1}(X_{2k+1} - c - dX_k)^2.$$

Consequently, we obviously have for all $n \geq 1$

$$(\widehat{\boldsymbol{\theta}}_n) = \begin{pmatrix} \widehat{a}_n \\ \widehat{b}_n \\ \widehat{c}_n \\ \widehat{d}_n \end{pmatrix} = \boldsymbol{\Sigma}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} X_{2k} \\ \delta_{2k} X_k X_{2k} \\ \delta_{2k+1} X_{2k+1} \\ \delta_{2k+1} X_k X_{2k+1} \end{pmatrix}, \quad (3.1)$$

where, for all $n \geq 0$,

$$\boldsymbol{\Sigma}_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix}, \quad \text{and} \quad \mathbf{S}_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix},$$

for $i \in \{0, 1\}$. In order to avoid intricate invertibility assumption, we shall assume, without loss of generality, that for all $n \geq 0$, $\boldsymbol{\Sigma}_n$ is invertible. Otherwise, we only have to add the identity matrix \mathbf{I}_4 to $\boldsymbol{\Sigma}_n$, as Proposition 4.2 states that the normalized limit of $\boldsymbol{\Sigma}_n$ is positive definite.

Remark 3.1. Note that when all data are observed, that is when all δ_k equal 1, this is simply the least squares estimator described in the previous literature. However, one must be careful here with the indices in the normalizing matrix, as there are now two different matrices \mathbf{S}_n^0 and \mathbf{S}_n^1 , while there was only one in the fully observed problem. The intuitive way to deal with missing data would be to restrict the sums to the observed data only. Note that our estimator is more complex as it involves sums depending on the absence or presence of even- or odd-type *daughters* of the available data.

We now turn to the estimation of the parameters σ^2 and ρ . We propose to estimate the conditional variance σ^2 and the conditional covariance ρ by

$$\widehat{\sigma}_n^2 = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}^*} (\widehat{\varepsilon}_{2k}^2 + \widehat{\varepsilon}_{2k+1}^2), \quad \widehat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1},$$

where for all $k \in \mathbb{G}_n$,

$$\begin{cases} \widehat{\varepsilon}_{2k} & = & \delta_{2k}(X_{2k} - \widehat{a}_n - \widehat{b}_n X_k), \\ \widehat{\varepsilon}_{2k+1} & = & \delta_{2k+1}(X_{2k+1} - \widehat{c}_n - \widehat{d}_n X_k). \end{cases},$$

and

$$\mathbb{T}_n^{*01} = \{k \in \mathbb{T}_n : \delta_{2k} \delta_{2k+1} = 1\},$$

so to speak \mathbb{T}_{n-1}^{*01} is the set of the cells of the tree \mathbb{T}_{n-1} which have exactly two offspring.

3.2. Main results

We can now state the sharp convergence results we obtain for the estimators above. We introduce additional notation. For $i \in \{0, 1\}$, let us denote:

$$\mathbf{L}^i = \begin{pmatrix} \pi z^i & h^i \\ h^i & k^i \end{pmatrix} \quad \mathbf{L}^{0,1} = \begin{pmatrix} \bar{p}(1, 1) & h^{0,1} \\ h^{0,1} & k^{0,1} \end{pmatrix}$$

where $\mathbf{z} = (z^0, z^1)$ is the left eigenvector for the dominant eigenvalue π of the descendants matrix \mathbf{P} introduced in section 2.2.2, h^i, k^i are defined in Propositions 6.3 and 6.5 and the four terms of $\mathbf{L}^{0,1}$ defined in Proposition 6.6. We also define the 4×4 matrices

$$\mathbf{\Sigma} = \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix}, \quad \text{and} \quad \mathbf{\Gamma} = \begin{pmatrix} \sigma^2 \mathbf{L}^0 & \rho \mathbf{L}^{0,1} \\ \rho \mathbf{L}^{0,1} & \sigma^2 \mathbf{L}^1 \end{pmatrix}. \tag{3.2}$$

Our first result deals with the strong consistency of the LS estimator $\hat{\boldsymbol{\theta}}_n$.

Theorem 3.2. *Under assumptions (HN.1), (HN.2), (HO) and (HI), $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}$ almost surely on $\bar{\mathcal{E}}$ with the rate of convergence*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.3}$$

In addition, we also have the quadratic strong law

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\hat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \mathbf{\Sigma} (\hat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) = 4 \frac{\pi - 1}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.4}$$

Our second result is devoted to the almost sure asymptotic properties of the variance and covariance estimators $\hat{\sigma}_n^2$ and $\hat{\rho}_n$. Let

$$\sigma_n^2 = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}^*} (\delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{2k+1}^2), \quad \rho_n = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k} \varepsilon_{2k} \delta_{2k+1} \varepsilon_{2k+1}.$$

Theorem 3.3. *Under assumptions (HN.1), (HN.2), (HO) and (HI), $\hat{\sigma}_n^2$ converges almost surely to σ^2 on $\bar{\mathcal{E}}$. More precisely, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \sum_{i=0}^1 \delta_{2k+i} (\hat{\varepsilon}_{2k+i} - \varepsilon_{2k+i})^2 = 4(\pi - 1) \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.5}$$

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\hat{\sigma}_n^2 - \sigma_n^2) = 4(\pi - 1) \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.6}$$

In addition, $\hat{\rho}_n$ converges almost surely to ρ on $\bar{\mathcal{E}}$ and one has

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (\hat{\varepsilon}_{2k} - \varepsilon_{2k}) \delta_{2k+1} (\hat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \\ = \rho \frac{\pi - 1}{\pi} \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned} \tag{3.7}$$

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\hat{\rho}_n - \rho_n) = \rho \frac{\pi - 1}{\bar{p}(1, 1)} \text{tr}((\mathbf{L}^1)^{-1}(\mathbf{L}^{0,1})^2(\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{3.8}$$

Our third result concerns the asymptotic normality for all our estimators $\hat{\boldsymbol{\theta}}_n, \hat{\sigma}_n^2$ and $\hat{\rho}_n$ given the non-extinction of the underlying Galton-Watson process. For this, using the fact that $\mathbb{P}(\bar{\mathcal{E}}) \neq 0$ thanks to Eq. (2.6), we define the probability $\mathbb{P}_{\bar{\mathcal{E}}}$ by

$$\mathbb{P}_{\bar{\mathcal{E}}}(A) = \frac{\mathbb{P}(A \cap \bar{\mathcal{E}})}{\mathbb{P}(\bar{\mathcal{E}})} \quad \text{for all } A \in \mathcal{A}.$$

Theorem 3.4. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have the central limit theorem*

$$\sqrt{|\mathbb{T}_{n-1}^*|}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Sigma}^{-1} \boldsymbol{\Gamma} \boldsymbol{\Sigma}^{-1}) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}). \tag{3.9}$$

In addition, we also have

$$\sqrt{|\mathbb{T}_n^*|}(\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0, \frac{\pi(\tau^4 - \sigma^4) + 2\bar{p}(1, 1)(\nu^2\tau^4 - \sigma^4)}{\pi}\right) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}), \tag{3.10}$$

where $\bar{p}(1, 1)$ is defined in Eq. (6.6) and

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|}(\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2\tau^4 - \rho^2) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}). \tag{3.11}$$

The proof of our main results is going to be detailed in the next sections. It is based on martingale properties, and we will exhibit our main martingale (\mathbf{M}_n) in Section 4. Sections 5 to 7 are devoted proving to the sharp asymptotic properties of (\mathbf{M}_n) . Finally, in Section 8 we prove our main results. Before turning to the definition of the martingale (\mathbf{M}_n) , we present a short application of our estimation procedure on data.

3.3. Results on real data

The biological issue addressed by Stewart et al. in [15] is aging in the single cell organism *Escherichia coli*, see also [7] for further biological details. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell has thus a new end (or *pole*), and an older one. The cell that inherits the old pole of its mother is called the old pole cell, the cell that inherits the new pole of its mother is called the new pole cell. Therefore, each cell has a *type*: old pole (even) or new pole (odd) cell, inducing asymmetry in the cell division.

Stewart et al. filmed colonies of dividing cells, determining the complete lineage and the growth rate of each cell. Their statistical study of the averaged genealogy and pair-wise comparison of sister cells showed that the old pole cells exhibit cumulatively slowed growth, less offspring biomass production and an increased probability of death. Note that their test assumes independence between the averaged pairs of sister cells which is not verified in the lineage.

TABLE 1
Estimation on the data set penna-2002-10-04-4

parameter	<i>a</i>	<i>c</i>
estimation	0.03627	0.03058
C.I.	[0.03276; 0.03979]	[0.02696; 0.03420]
parameter	<i>b</i>	<i>d</i>
estimation	0.02662	0.17055
C.I.	[−0.06866; 0.12191]	[0.07247; 0.26863]

Another analysis was proposed in [9]. They model the growth rate by a Markovian bifurcating process, allowing single-experiment statistical analysis instead of averaging all the genealogical trees. Asymptotic properties of a more general asymmetric Markovian bifurcating autoregressive process are then investigated in [8], where a Wald's type test is rigorously constructed to study the asymmetry of the process. These results cannot be compared to ours because this model does not take into account the possibly missing data from the genealogies, and it is not clear how the author manages them, as not a single tree from the data of [15] is complete. In [5], the authors take missing data into account but, contrary to our approach, they allow different sets of parameters for cells with two, one or no offspring, making the direct comparison with our estimator again impossible.

We have applied our methodology on the set of data `penna-2002-10-04-4` from the experiments of [15]. It is the largest data set of the experiment. It contains 663 cells up to generation 9 (note that there would be 1023 cells in a full tree up to generation 9). In particular, we have performed

- point estimation of the vector θ ,
- interval estimation for the coefficients (a, b, c, d) ,
- Wald's type symmetry tests for the entries of $\hat{\theta}_n$.

Table 1 gives the estimation $\hat{\theta}_9$ of θ with the 95% Confidence Interval (C.I.) of each coefficient. The variance given by the CLT for θ in Eq. (3.9), is approximated by $\Sigma_n^{-1} \Gamma_n \Sigma_n^{-1}$ thanks to the convergence given in Corollary 4.3. The confidence intervals of b and d show that the *non explosion* assumption ($|b| < 1$ and $|d| < 1$) is satisfied. Some empirical computation on the process (δ_k) gives the following estimation for the highest eigenvalue of the Galton-Watson process: $\hat{\pi} = 1.35669$ (with confidence interval [1.27979, 1.43361], see [14]), also satisfying the super-criticality assumption. Wald tests of comparison between the coefficients of θ have been deduced of the CLT. The null hypotheses $(a, b) = (c, d)$ (resp. $a = c, b = d$) are rejected with p-values $p = 0.0211$ (resp. $p = 0.0158$ and $p = 0.0244$). Hence on this data set the cell division is indeed statistically asymmetric.

4. Martingale approach

To establish all the asymptotic properties of our estimators, we shall make use of a martingale approach, similar to [3]. However, their results cannot be

used in our framework, since the randomness comes now not only from the state process, but also from the time space (genealogy). These two mixed randomness sources require careful choice between various filtrations, and stronger moment assumptions on the driven noise of the BAR process. For all $n \geq 1$, denote

$$M_n = \sum_{k \in \mathbb{T}_{n-1}} (\delta_{2k}\varepsilon_{2k}, \delta_{2k}X_k\varepsilon_{2k}, \delta_{2k+1}\varepsilon_{2k+1}, \delta_{2k+1}X_k\varepsilon_{2k+1})^t.$$

Thus, for all $n \geq 2$, we readily deduce from Equations (3.1) and (2.1) that

$$\hat{\theta}_n - \theta = \Sigma_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k}\varepsilon_{2k} \\ \delta_{2k}X_k\varepsilon_{2k} \\ \delta_{2k+1}\varepsilon_{2k+1} \\ \delta_{2k+1}X_k\varepsilon_{2k+1} \end{pmatrix} = \Sigma_{n-1}^{-1} M_n. \tag{4.1}$$

The key point of our approach is that (M_n) is a martingale for a well chosen filtration.

4.1. Martingale property

Recall that $\mathcal{O} = \sigma\{\delta_k, k \in \mathbb{T}\}$ is the σ -field generated by the observation process. We shall assume that all the history of the process (δ_k) is known at time 0 and use the filtration $\mathbb{F}^{\mathcal{O}} = (\mathcal{F}_n^{\mathcal{O}})$ defined for all n by

$$\mathcal{F}_n^{\mathcal{O}} = \mathcal{O} \vee \sigma\{\delta_k X_k, k \in \mathbb{T}_n\} = \mathcal{O} \vee \sigma\{X_k, k \in \mathbb{T}_n^*\},$$

where $\mathcal{F} \vee \mathcal{G}$ denotes the σ -field generated by both \mathcal{F} and \mathcal{G} . Note that for all n , $\mathcal{F}_n^{\mathcal{O}}$ is a sub σ -field of $\mathcal{O} \vee \mathcal{F}_n$.

Proposition 4.1. *Under assumptions (HN.1), (HN.2) and (HI), the process (M_n) is a square integrable $\mathbb{F}^{\mathcal{O}}$ -martingale with increasing process given, for all $n \geq 1$, by*

$$\langle M \rangle_n = \Gamma_{n-1} = \begin{pmatrix} \sigma^2 \mathbf{S}_{n-1}^0 & \rho \mathbf{S}_{n-1}^{0,1} \\ \rho \mathbf{S}_{n-1}^{0,1} & \sigma^2 \mathbf{S}_{n-1}^1 \end{pmatrix},$$

where \mathbf{S}_n^0 and \mathbf{S}_n^1 are defined in section 3.1 and

$$\mathbf{S}_n^{0,1} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

Proof. First, notice that for all $n \geq 1$, one has

$$\Delta M_n = M_n - M_{n-1} = \sum_{k \in \mathbb{G}_{n-1}} \begin{pmatrix} \delta_{2k}\varepsilon_{2k} \\ \delta_{2k}X_k\varepsilon_{2k} \\ \delta_{2k+1}\varepsilon_{2k+1} \\ \delta_{2k+1}X_k\varepsilon_{2k+1} \end{pmatrix}.$$

Now, we use the fact that for all n , $\mathcal{F}_n^\mathcal{O}$ is a sub- σ field of $\mathcal{O} \vee \mathcal{F}_n$, the independence between \mathcal{O} and \mathcal{F}_n under assumption **(HI)** and the moment hypothesis **(HN.1)** to obtain

$$\begin{aligned} \mathbb{E}[\delta_{2k}\varepsilon_{2k} \mid \mathcal{F}_{n-1}^\mathcal{O}] &= \delta_{2k}\mathbb{E}[\mathbb{E}[\varepsilon_{2k} \mid \mathcal{O} \vee \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}^\mathcal{O}] \\ &= \delta_{2k}\mathbb{E}[\mathbb{E}[\varepsilon_{2k} \mid \mathcal{F}_{n-1}] \mid \mathcal{F}_{n-1}^\mathcal{O}] = 0. \end{aligned}$$

We obtain similar results for the other entries of ΔM_n as δ_{2k+1} and X_k are also $\mathcal{F}_{n-1}^\mathcal{O}$ -measurable. Hence, (M_n) is a $\mathbb{F}^\mathcal{O}$ -martingale. It is clearly square integrable from assumption **(HN.1)**. The same measurability arguments together with assumption **(HN.2)** yield

$$\begin{aligned} &\mathbb{E}[\Delta M_n(\Delta M_n)^t \mid \mathcal{F}_{n-1}^\mathcal{O}] \\ &= \sum_{k \in \mathbb{G}_{n-1}} \begin{pmatrix} \sigma^2 \delta_{2k} & \sigma^2 \delta_{2k} X_k & \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k \\ \sigma^2 \delta_{2k} X_k & \sigma^2 \delta_{2k} X_k^2 & \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 \\ \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} & \sigma^2 \delta_{2k+1} X_k \\ \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 & \sigma^2 \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} X_k^2 \end{pmatrix}. \end{aligned}$$

Hence the result as $\langle M \rangle_n = \sum_{\ell=1}^n \mathbb{E}[\Delta M_\ell(\Delta M_\ell)^t \mid \mathcal{F}_{\ell-1}^\mathcal{O}]$. □

Our main results are direct consequences of the sharp asymptotic properties of the martingale (M_n) . In particular, we will extensively use the strong law of large numbers for locally square integrable real martingales given in Theorem 1.3.15 of [6]. Throughout this paper, we shall also use other auxiliary martingales, either with respect to the same filtration $\mathbb{F}^\mathcal{O}$, or with respect to other filtrations naturally embedded in our process, see Lemma 5.1.

4.2. Asymptotic results

We first give the asymptotic behavior of the matrices S_n^0 , S_n^1 and $S_n^{0,1}$. This is the first step of our asymptotic results.

Proposition 4.2. *Suppose that assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)** are satisfied. Then, for $i \in \{0, 1\}$, we have*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{S_n^i}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\mathcal{E}}} L^i \quad a.s. \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{S_n^{0,1}}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\mathcal{E}}} L^{0,1} \quad a.s.$$

In addition, L^0 and L^1 , hence Σ are definite positive.

A consequence of this proposition is the asymptotic behavior of the increasing process of the martingale (M_n) .

Corollary 4.3. *Suppose that assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)** are satisfied. Then, we have*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Sigma_n}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\mathcal{E}}} \Sigma, \quad \text{and} \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Gamma_n}{|\mathbb{T}_n^*|} = \mathbb{1}_{\bar{\mathcal{E}}} \Gamma.$$

This result is the keystone of our asymptotic analysis. It enables us to prove sharp asymptotic properties for the martingale (\mathbf{M}_n) .

Theorem 4.4. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathbf{M}_n^t \Sigma_{n-1}^{-1} \mathbf{M}_n = \mathcal{O}(n) \quad a.s. \tag{4.2}$$

In addition, we also have

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \Sigma_{\ell-1}^{-1} \mathbf{M}_\ell = 4\sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \tag{4.3}$$

Moreover, we have the central limit theorem on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} \mathbf{M}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}).$$

As seen in Eq. (4.1), $(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$ is closely linked to \mathbf{M}_n and this last theorem is then the major step to establish the asymptotic properties of our estimators. The proof of this Theorem is given in Section 7. As explained before, it is a consequence of Proposition 4.2 which proof is detailed in Section 6. In between, Section 5 presents preliminary results in the form of laws of large number for the observation, noise and BAR processes.

5. Laws of large numbers

We now state some laws of large numbers involving the observation, noise and BAR processes. They are based on martingale convergence results, and we start with giving a general result of convergence for martingales adapted to our framework.

5.1. Martingale convergence results

The following result is nothing but the strong law of large numbers for square integrable martingales, written in our peculiar setting, and will be repeatedly used.

Lemma 5.1. *Let $\mathcal{G} = (\mathcal{G}_n)$ be some filtration, (H_n) and (G_n) be two sequences of random variables satisfying the following hypotheses:*

- (i) *for all $n \geq 1$, for all $k \in \mathbb{G}_n$, H_k is \mathcal{G}_{n-1} -measurable, G_k is \mathcal{G}_n -measurable, and $\mathbb{E}[(H_k G_k)^2] < +\infty$,*
- (ii) *there exist $c^2 > 0$, $r \in [-1, 1]$, such that for all $n \geq 1$, for all $k, p \in \mathbb{G}_n$,*

$$\mathbb{E}[G_k | \mathcal{G}_{n-1}] = 0, \quad \mathbb{E}[G_k G_p | \mathcal{G}_{n-1}] = \begin{cases} c^2 & \text{if } k = p, \\ rc^2 & \text{if } k \neq p \text{ and } [k/2] = [p/2], \\ 0 & \text{otherwise,} \end{cases}$$

(iii) there exists a sequence of real numbers (a_n) that tends to ∞ such that $\sum_{k \in \mathbb{T}_n} H_k^2 = \mathcal{O}(a_n)$.

Then $\sum_{k \in \mathbb{T}_n} H_k G_k$ is a \mathcal{G} -martingale and one has

$$\lim_{n \rightarrow \infty} \frac{1}{a_n} \sum_{k \in \mathbb{T}_n} H_k G_k = 0 \quad a.s.$$

Proof. Define $D_n = \sum_{k \in \mathbb{T}_n} H_k G_k$. Assumptions (i) and (ii) clearly yield that (D_n) is a square integrable martingale with respect to the filtration (\mathcal{G}_n) . Thanks to (ii), its increasing process satisfies

$$\begin{aligned} \langle D \rangle_n &= c^2 \left(\sum_{k \in \mathbb{T}_n} H_k^2 + 2r \sum_{k \in \mathbb{T}_{n-1}} H_{2k} H_{2k+1} \right) \\ &\leq c^2 \left(\sum_{k \in \mathbb{T}_n} H_k^2 + r \sum_{k \in \mathbb{T}_{n-1}} (H_{2k}^2 + H_{2k+1}^2) \right) \\ &\leq c^2 (r + 1) \sum_{k \in \mathbb{T}_n} H_k^2, \end{aligned}$$

and now, (iii) implies that $\langle D \rangle_n = \mathcal{O}(a_n)$. Finally, since the sequence (a_n) tends to ∞ , Theorem 1.3.15 of [6] ensures that $D_n = o(a_n)$ a.s. \square

We also recall Lemma A.3 of [3] that will be useful in the sequel.

Lemma 5.2. *Let (A_n) be a sequence of real-valued matrices such that*

$$\sum_{n=0}^{\infty} \|A_n\| < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \sum_{k=0}^n A_k = A.$$

In addition, let (X_n) be a sequence of real-valued vectors which converges to a limiting value X . Then,

$$\lim_{n \rightarrow \infty} \sum_{\ell=0}^n A_{n-\ell} X_\ell = AX.$$

5.2. Laws of large numbers for the observation process

We now give more specific results on the asymptotic behavior of the observation process $(\delta_k)_{k \geq 1}$. Recall the notation \mathbb{T}_n^i defined in (2.2).

Lemma 5.3. *Under the assumption (HO), we have the following convergences, for (i, j) in $\{0, 1\}^2$*

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_{2k+j} = p_{ij} \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_{2k} \delta_{2k+1} = p^{(i)}(1, 1) \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

Proof. Recall that $\delta_{2k+j} = \delta_k \zeta_k^j$, so that

$$\sum_{k \in \mathbb{T}_n^i} \delta_{2k+j} = p_{ij} \sum_{k \in \mathbb{T}_n^i} \delta_k + \sum_{k \in \mathbb{T}_n^i} \delta_k (\zeta_k^j - p_{ij}) = p_{ij} \left(i + \sum_{\ell=1}^n Z_\ell^i \right) + D_n,$$

since $\mathbb{G}_0 = \{1\}$, so that \mathbb{T}_n^i contains 1 or not, according to $i = 1$ or not, and where $D_n = \sum_{k \in \mathbb{T}_n^i} \delta_k (\zeta_k^j - p_{ij})$. To deal with D_n , we use Lemma 5.1, with $\mathcal{G} = (\mathcal{Z}_n)$ (recall that $\mathcal{Z}_n = \sigma\{\zeta_k, k \in \mathbb{T}_n\}$), $H_k = \delta_k \mathbb{1}_{\{k \in \mathbb{T}^i\}}$, and $G_k = (\zeta_k^j - p_{ij}) \mathbb{1}_{\{k \in \mathbb{T}^i\}}$. Assumption (i) of Lemma 5.1 is obviously satisfied, since δ_k , for $k \in \mathbb{G}_n$, is \mathcal{Z}_{n-1} -measurable. Regarding (ii), since the sequence (ζ_k^j) is a sequence of i.i.d. random variables with expectation p_{ij} and variance σ_{ij}^2 , we have $\mathbb{E}[G_k | \mathcal{Z}_{n-1}] = 0$ and $\mathbb{E}[G_k^2 | \mathcal{Z}_{n-1}] = \sigma_{ij}^2$, for $k \in \mathbb{G}_n$, and $\mathbb{E}[G_k G_p | \mathcal{Z}_{n-1}] = 0$, for $k \neq p \in \mathbb{G}_n$. Finally, we turn to assumption (iii):

$$\sum_{k \in \mathbb{T}_n} H_k^2 = \sum_{k \in \mathbb{T}_n^i} \delta_k = i + \sum_{\ell=1}^n Z_\ell^i = \mathcal{O}(\pi^n),$$

thanks to (HO) and Eq. (2.7). Finally, $D_n = o(\pi^n)$, and again using Eq. (2.7), we obtain the first limit. The proof of the second one is similar using the \mathcal{Z} -martingale:

$$\sum_{k \in \mathbb{T}_n^i} \delta_k (\delta_{2k} \delta_{2k+1} - p^{(i)}(1, 1)) = \sum_{k \in \mathbb{T}_n} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_k}_{H_k} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} (\zeta_k^0 \zeta_k^1 - p^{(i)}(1, 1))}_{G_k},$$

and Lemma 5.1 again. □

5.3. Laws of large numbers for the noise process

We need to establish strong laws of large numbers for the noise sequence (ε_n) restricted to the observed indices.

Lemma 5.4. *Under assumptions (HN.1), (HN.2), (HO), (HI) and for $i \in \{0, 1\}$, one has*

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k+i} \varepsilon_{2k+i} = 0 \quad a.s.$$

Proof. Set

$$P_n^i = \sum_{k \in \mathbb{T}_n} \underbrace{\delta_{2k+i}}_{H_k} \underbrace{\varepsilon_{2k+i}}_{G_k}.$$

We use Lemma 5.1, with $\mathcal{G} = (\mathbb{F}_{n+1}^O)$. Assumption (i) is obvious. For $k \in \mathbb{G}_{n+1}^i$, we have $\mathbb{E}[G_k | \mathbb{F}_{n+1}^O] = 0$ and $\mathbb{E}[G_k^2 | \mathbb{F}_{n+1}^O] = \sigma^2$, and $\mathbb{E}[G_k G_p | \mathbb{F}_{n+1}^O] = 0$, for $k \neq p \in \mathbb{G}_{n+1}^i$. Finally, we turn to assumption (iii):

$$\sum_{k \in \mathbb{T}_n} H_k^2 = \sum_{k \in \mathbb{T}_n} \delta_{2k+i}^2 = \sum_{\ell=1}^{n+1} Z_\ell^i = \mathcal{O}(\pi^n),$$

as n tends to infinity, thanks to Eq. (2.7). We obtain the result. \square

Lemma 5.5. *Under assumptions (HN.1), (HN.2), (HO), (HI) and for $i \in \{0, 1\}$, one has*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_k &= \sigma^2 \frac{\pi}{\pi - 1} Wz^i \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} &= \rho p^{(i)}(1, 1) \frac{\pi}{\pi - 1} Wz^i \quad a.s. \end{aligned}$$

Proof. In order to prove the first convergence, we apply again Lemma 5.1 to the $\mathbb{F}^{\mathcal{O}}$ -martingale:

$$Q_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} (\varepsilon_k^2 - \sigma^2) \delta_k = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_k}_{H_k} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} (\varepsilon_k^2 - \sigma^2)}_{G_k},$$

Under (HN.1), (HN.2), we have $\mathbb{E}[G_k | \mathbb{F}_n^{\mathcal{O}}] = 0$ and $\mathbb{E}[G_k^2 | \mathbb{F}_n^{\mathcal{O}}] = \tau^4 - \sigma^4$, and $\mathbb{E}[G_k G_p | \mathbb{F}_n^{\mathcal{O}}] = 0$, for $k \neq p \in \mathbb{G}_n$. Thanks to Eq. (2.7), we have:

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i} \delta_k = \frac{1}{\pi^n} \sum_{\ell=1}^n Z_\ell^i \xrightarrow[n \rightarrow \infty]{} \frac{\pi}{\pi - 1} Wz^i \quad a.s.$$

which both implies assumption (iii) and the first convergence. To prove the second convergence, we write

$$\begin{aligned} &\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} \\ &= \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} \delta_{2k} \delta_{2k+1}}_{H_k} \underbrace{\mathbb{1}_{\{k \in \mathbb{T}^i\}} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho)}_{G_k} + \frac{1}{\pi^n} \rho \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \end{aligned}$$

We use Lemma 5.1 to prove that the first term converges to 0; Lemma 5.3 gives the limit of the second term. \square

Corollary 5.6. *Under assumptions (HN.1), (HN.2), (HO), (HI) and for $i \in \{0, 1\}$, one has*

$$\begin{aligned} \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+j} &= \sigma^2 p_{ij} \frac{\pi}{\pi - 1} Wz^i \quad a.s. \\ \lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k} \delta_{2k+1} \varepsilon_{2k} \varepsilon_{2k+1} &= \rho \bar{p}(1, 1) \frac{\pi}{\pi - 1} W \quad a.s. \end{aligned}$$

Proof. The proof of the first limit is similar to the preceding ones, using the decomposition $\delta_{2k+j} = \delta_k \zeta_k^j$ and the properties of the sequence (ζ_n^j) . Using Lemma 5.5 the second one is straightforward. \square

Lemma 5.7. *Under assumptions (HN.1), (HN.2), (HO), (HI) and for $i \in \{0, 1\}$, we have*

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_k \varepsilon_k^4 = \tau^4 \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

$$\lim_{n \rightarrow +\infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_{n-1}^i} \delta_{2k} \delta_{2k+1} \varepsilon_{2k}^2 \varepsilon_{2k+1}^2 = \nu^2 \tau^4 p^{(i)}(1, 1) \frac{\pi}{\pi - 1} W z^i \quad a.s.$$

Proof. The proof follows essentially the same lines as the proof of Lemma 5.5 using the square integrable real martingales

$$Q_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_k (\varepsilon_k^4 - \tau^4), \quad \text{and} \quad R_n = \sum_{k \in \mathbb{T}_n^i \setminus \mathbb{T}_0} \delta_{2j} \delta_{2j+1} (\varepsilon_{2j}^2 \varepsilon_{2j+1}^2 - \nu^2 \tau^4).$$

It is therefore left to the reader. □

6. Convergence of the increasing process

We can now turn to the proof of our keystone result, the convergence of the increasing process of the main martingale (M_n) .

6.1. Preliminary results

We first need an upper bound of the normalized sums of the $\delta_{2n+i} X_n^2$, and $\delta_{2n} \delta_{2n+1} X_n^2$ before being able to deduce their limits.

Lemma 6.1. *Under assumptions (HN.1), (HN.2), (HI) and (HO), and for $i \in \{0, 1\}$, we have*

$$\sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^2 = \mathcal{O}(\pi^n) \quad \text{and} \quad \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2 = \mathcal{O}(\pi^n) \quad a.s.$$

Proof. In all the sequel, for all $k \geq 1$, define $a_{2k} = a$, $b_{2k} = b$, $a_{2k+1} = c$, $b_{2k+1} = d$ and $\eta_k = a_k + \varepsilon_k$ with the convention that $\eta_1 = 0$. It follows from a recursive application of relation (2.1) that, for all $k \geq 1$,

$$X_k = \left(\prod_{\ell=0}^{r_k-1} b_{\lfloor \frac{k}{2^\ell} \rfloor} \right) X_1 + \sum_{\ell=0}^{r_k-1} \left(\prod_{p=0}^{\ell-1} b_{\lfloor \frac{k}{2^p} \rfloor} \right) \eta_{\lfloor \frac{k}{2^\ell} \rfloor}, \tag{6.1}$$

with the convention that an empty product equals 1. Set $\alpha = \max(|a|, |c|)$, $\beta = \max(|b|, |d|)$ and notice that $0 < \beta < 1$. The proof of Lemma A.5 in [3] yields

$$\begin{aligned}
 \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} X_k^2 &\leq \frac{4}{1-\beta} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{\lfloor \frac{k}{2^\ell} \rfloor}^2 + \frac{4\alpha^2}{1-\beta} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \\
 &\quad + 2X_1^2 \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{2r_k}, \\
 &\leq \frac{4}{1-\beta} A_n^i + \frac{4\alpha^2}{1-\beta} B_n^i + 2X_1^2 C_n^i, \tag{6.2}
 \end{aligned}$$

where, for $i \in \{0, 1\}$,

$$A_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{\lfloor \frac{k}{2^\ell} \rfloor}^2, \quad B_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell, \quad C_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{2r_k}.$$

The last two terms above are readily evaluated by splitting the sums generation-wise. Indeed, the last term can be rewritten as

$$C_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \beta^{2\ell} = \sum_{\ell=1}^n \beta^{2\ell} Z_{\ell+1}^i = \pi^n \sum_{\ell=1}^n (\pi^{-1})^{n-\ell} \left(\beta^{2\ell} \frac{Z_{\ell+1}^i}{\pi^\ell} \right).$$

We now use Lemma 5.2 with $A_n = \pi^{-n}$ and $X_n = \beta^{2n} Z_{n+1}^i \pi^{-n}$. On the one hand, the series of (π^{-n}) converges to $\pi/(\pi - 1)$ as $\pi > 1$ by assumption; on the other hand, β^{2n} tends to 0 as n tends to infinity as $\beta < 1$, and $Z_n^i \pi^{-n}$ converges a.s. to Wz^i according to Eq. (2.7), hence $\beta^{2n} Z_{n+1}^i \pi^{-n}$ tends to 0 as n tends to infinity. Lemma 5.2 thus yields

$$\lim_{n \rightarrow \infty} \sum_{\ell=1}^n (\pi^{-1})^{n-\ell} \left(\beta^{2\ell} \frac{Z_{\ell+1}^i}{\pi^\ell} \right) = 0 \quad \text{and} \quad C_n^i = o(\pi^n) \quad \text{a.s.}$$

We now turn to the term B_n^i :

$$B_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \frac{1-\beta^\ell}{1-\beta} \leq \frac{1}{(1-\beta)} \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \leq \frac{|\mathbb{T}_{n+1}^*|}{(1-\beta)} = \mathcal{O}(\pi^n),$$

due to Lemma 2.1. It remains to control the first term A_n^i . Note that ε_k appears in A_n^i as many times as it has descendants up to the n -th generation, and its multiplicative factor for its p -th generation descendant k is $\beta^p \delta_{2k}$. This leads to

$$A_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \varepsilon_k^2 \sum_{p=0}^{n-\ell} \beta^p \sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i}.$$

Now, note that $\sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i} = \delta_k \sum_{m=0}^{2^p-1} \delta_{2(2^p k+m)+i}$ is the number of descendants of type i of individual k after $p + 1$ generations. We denote it $Z_{p+1}^i(k)$, and split A_n^i the following way:

$$A_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \sigma^2 \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) + \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k). \tag{6.3}$$

We first deal with the second term of the above sum.

$$\begin{aligned} \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) &= \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \delta_k Z_{p+1}^i(k) \\ &= \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Y_{\ell,p}^i, \end{aligned}$$

where $Y_{\ell,p}^i = \sum_{k \in \mathbb{G}_\ell} (\varepsilon_k^2 - \sigma^2) \delta_k Z_{p+1}^i(k)$. Tedious but straightforward computations lead to the following expression for the second order moment of $Y_{\ell,p}^i$, relying on assumptions **(HI)**, **(HN.1)** and **(HN.2)**. We also use the fact that, for $k \in \mathbb{G}_\ell$, conditionally to $\{\delta_k = 1\}$, $Z_{p+1}^i(k)$ follows the same law as Z_{p+1}^i , and is independent of any $Z_{p+1}^i(k')$, for $k' \neq k \in \mathbb{G}_\ell$.

$$\begin{aligned} \mathbb{E}[(Y_{\ell,p}^i)^2] &= (\tau^4 - \sigma^4) \mathbb{E}[Z_\ell^0 + Z_\ell^1] \mathbb{E}[(Z_{p+1}^i)^2] \\ &\quad + (\nu^2 \tau^4 - \sigma^4) \mathbb{E}[Z_{p+1}^i]^2 \mathbb{E}\left[\sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k} \delta_{2k+1}\right] \\ &\leq (\tau^4 - \sigma^4) \mathbb{E}[Z_\ell^0 + Z_\ell^1] \left(\mathbb{E}[(Z_{p+1}^i)^2] + \mathbb{E}[Z_{p+1}^i]^2\right), \end{aligned}$$

since $\sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k} \delta_{2k+1} \leq \sum_{k \in \mathbb{G}_{\ell-1}} (\delta_{2k} + \delta_{2k+1}) = Z_\ell^0 + Z_\ell^1$. Now, using results on the moments of a two-type Galton-Watson process (see e.g. [12]), we know that $\mathbb{E}[(Z_{p+1}^i)^2] = \mathcal{O}(\pi^{2p})$. Recall Eq. (2.7) to obtain that $\mathbb{E}[(Y_{\ell,p}^i)^2] = \mathcal{O}(\pi^\ell \pi^{2p})$, which immediately entails that $|Y_{\ell,p}^i| = o(\pi^{\alpha\ell} \pi^{\gamma p})$ a.s., for any $\alpha > 1/2$ and $\gamma > 1$. We thus one gets

$$\sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Y_{\ell,p}^i = \mathcal{O}((\beta\pi^\gamma)^n) = \mathcal{O}(\pi^n) \quad \text{a.s.},$$

since we can choose γ close enough to 1 to get $\beta\pi^\gamma \leq \pi$, as $\beta < 1$. We have thus proved that the second term in the sum in (6.3) is $\mathcal{O}(\pi^n)$, we now turn to the first one

$$\begin{aligned} &\sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \sigma^2 \sum_{p=0}^{n-\ell} \beta^p \delta_k Z_{p+1}^i(k) \\ &= \sigma^2 \sum_{\ell=1}^n \sum_{p=0}^{n-\ell} \beta^p \sum_{k \in \mathbb{G}_\ell} \delta_k Z_{p+1}^i(k) = \sigma^2 \sum_{\ell=1}^n \sum_{p=0}^{n-\ell} \beta^p Z_{\ell+p+1}^i \\ &= \sigma^2 \sum_{p=0}^{n-1} \beta^p \sum_{\ell=1}^{n-p} Z_{\ell+p+1}^i \leq \sigma^2 \sum_{p=0}^{n-1} \beta^p |\mathbb{T}_{n+1}^*| = \mathcal{O}(\pi^n) \quad \text{a.s.} \end{aligned}$$

Finally, $A_n^i = \mathcal{O}(\pi^n)$, and the first result of the Lemma is proved. The second result follows immediately from the remark that the second sum in Lemma 6.1 is clearly smaller than the first one. \square

Lemma 6.2. *Under assumptions (HN.1), (HN.2), (HI) and (HO), and for $i \in \{0, 1\}$, we have*

$$\sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^4 = \mathcal{O}(\pi^n) \quad \text{and} \quad \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^4 = \mathcal{O}(\pi^n) \quad a.s.$$

Proof. The proof mimics that of Lemma 6.1. Instead of Equation (6.2), we have

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} X_k^4 \leq \frac{64}{(1-\beta)^3} A_n^i + \frac{64\alpha^4}{(1-\beta)^3} B_n^i + 8X_1^4 C_n^i$$

with, for i in $\{0, 1\}$

$$A_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell \varepsilon_{[\frac{k}{2^\ell}]}, \quad B_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \sum_{\ell=0}^{r_k-1} \beta^\ell, \quad C_n^i = \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i} \beta^{4r_k}.$$

We can easily prove that $(B_n^i + C_n^i) = \mathcal{O}(\pi^n)$. Therefore, we only need a sharper estimate for A_n^i . Via the same lines as in the proof of Lemma 6.1, but dealing with ε_k^4 instead of ε_k^2 , we can show that $A_n^i = \mathcal{O}(\pi^n)$ a.s. which immediately yields the first result. The second one is obtained by remarking that the second sum is less than the first one. \square

6.2. Asymptotic behavior of the sum of observed data

We now turn to the asymptotic behavior of the sums of the observed data. More precisely, set $H_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k$, for i in $\{0, 1\}$, and $\mathbf{H}_n = (H_n^0, H_n^1)^t$. The following result gives the asymptotic behavior of (\mathbf{H}_n) .

Proposition 6.3. *Under assumptions (HN.1), (HN.2), (HI) and (HO), we have the convergence:*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{H}_n}{\pi^n} = \frac{\pi}{\pi-1} W \mathbf{h} \quad a.s.,$$

where

$$\mathbf{h} = \begin{pmatrix} h^0 \\ h^1 \end{pmatrix} = (\mathbf{I}_2 - \tilde{\mathbf{P}}_1)^{-1} \mathbf{P}^t \begin{pmatrix} az^0 \\ cz^1 \end{pmatrix} \quad \text{and} \quad \tilde{\mathbf{P}}_1 = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b & 0 \\ 0 & d \end{pmatrix}.$$

Proof. We first prove that the sequence (\mathbf{H}_n) satisfies a recursive property using Equation (2.1).

$$\begin{aligned} H_n^0 &= X_1 \delta_2 + \sum_{k \in \mathbb{T}_n^0} \left(a + bX_{[\frac{k}{2}] + \varepsilon_k \right) \delta_{2k} + \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \left(c + dX_{[\frac{k}{2}] + \varepsilon_k \right) \delta_{2k} \\ &= X_1 \delta_2 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k} + b \sum_{k \in \mathbb{T}_n^0} X_{[\frac{k}{2}] \delta_{2k} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k} + d \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} X_{[\frac{k}{2}] \delta_{2k} \\ &\quad + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k} \\ &= bp_{00} H_{n-1}^0 + dp_{10} H_{n-1}^1 + B_n^0, \end{aligned}$$

with

$$\begin{aligned}
 B_n^0 &= X_1 \delta_2 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k} + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k} \\
 &\quad + b \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} (\delta_{4k} - p_{00}) + d \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} (\delta_{4k+2} - p_{10}).
 \end{aligned}$$

Similarly, we have

$$H_n^1 = bp_{01}H_{n-1}^0 + dp_{11}H_{n-1}^1 + B_n^1,$$

with

$$\begin{aligned}
 B_n^1 &= X_1 \delta_3 + a \sum_{k \in \mathbb{T}_n^0} \delta_{2k+1} + c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k+1} + \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+1} \\
 &\quad + b \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} (\delta_{4k+1} - p_{01}) + d \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} (\delta_{4k+3} - p_{11}).
 \end{aligned}$$

Let us denote $\mathbf{B}_n = (B_n^0, B_n^1)^t$. The last equations yield in the matrix form:

$$\frac{\mathbf{H}_n}{\pi^n} = \tilde{\mathbf{P}}_1 \frac{\mathbf{H}_{n-1}}{\pi^{n-1}} + \frac{\mathbf{B}_n}{\pi^n} = \tilde{\mathbf{P}}_1^n \mathbf{H}_0 + \sum_{k=1}^n \tilde{\mathbf{P}}_1^{n-k} \frac{\mathbf{B}_k}{\pi^k},$$

with

$$\tilde{\mathbf{P}}_1 = \frac{1}{\pi} \begin{pmatrix} bp_{00} & dp_{10} \\ bp_{01} & dp_{11} \end{pmatrix} = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b & 0 \\ 0 & d \end{pmatrix}.$$

One has $\|\tilde{\mathbf{P}}_1^n\| \leq \pi^{-n} \beta^n \|\mathbf{P}^n\|$. It is well known that $\pi^{-n} \mathbf{P}^n$ converges to a fixed matrix (see e.g. [13]) as \mathbf{P} is a positive matrix with dominant eigenvalue π . Since $\beta < 1$, the sequence $\tilde{\mathbf{P}}_1^n$ thus converges to 0 as n tends to infinity. In addition, $\sum \|\tilde{\mathbf{P}}_1^n\|$ is bounded, $\mathbf{I}_2 - \tilde{\mathbf{P}}_1$ is invertible and $\sum_{n \geq 0} \tilde{\mathbf{P}}_1^n$ converges to $(\mathbf{I}_2 - \tilde{\mathbf{P}}_1)^{-1}$. In order to use Lemma 5.2, we need to compute the limit of \mathbf{B}_n/π^n . First, we prove that

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+i} = o(\pi^n), \tag{6.4}$$

for $i \in \{0, 1\}$, thanks to Lemma 5.1. Indeed, set $\mathcal{G} = \mathbb{F}^{\mathcal{O}}$, $H_k = \delta_{2k+i}$, $G_k = \varepsilon_k$. Thus hypothesis (i) of Lemma 5.1 is obvious, (ii) comes from (HN.1) and (HN.2). Finally, the last assumption (iii) holds, since

$$\sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \delta_{2k+i}^2 = \sum_{\ell=1}^{n+1} Z_\ell^i = \mathcal{O}(\pi^n),$$

the last equality coming from (2.7), which holds thanks to (HO). Now, we turn to the terms

$$\sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = \sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\zeta_{2k+i}^j - p_{ij}),$$

for $(i, j) \in \{0, 1\}^2$. We use again Lemma 5.1, with the following setting: $(\mathcal{G}_n) = (\mathcal{Z}_{n+1} \vee \mathcal{F}_{n+1})$, $H_k = X_k \delta_{2k+i}$, $G_k = \zeta_{2k+i}^j - p_{ij}$. For $k \in \mathbb{T}_n$, we check that $X_k \delta_{2k+i}$ is \mathcal{G}_{n-1} -measurable, since X_k is \mathcal{F}_n -measurable and δ_{2k+i} is \mathcal{Z}_n -measurable. Next, because of (HI) and of the independence of the sequence (ζ_k) , $\mathbb{E}[\zeta_{2k+i}^j - p_{ij} | \mathcal{Z}_n \vee \mathcal{F}_n] = 0$. The same independence hypothesis yields that $\mathbb{E}[G_k G_p | \mathcal{Z}_n \vee \mathcal{F}_n] \neq 0$ only if $k = p$, and then equals σ_{ij}^2 . Finally,

$$\sum_{k \in \mathbb{T}_n} (X_k \delta_{2k+i})^2 = \sum_{k \in \mathbb{T}_n} X_k^2 \delta_{2k+i} = \mathcal{O}(\pi^n),$$

thanks to Lemma 6.1. Now, Lemma 5.1 allows to conclude that

$$\sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = o(\pi^n), \tag{6.5}$$

for $(i, j) \in \{0, 1\}^2$. Next, Lemma 5.3 gives the limit of the term $\sum_{k \in \mathbb{T}_n} \delta_{2k+j}$, for $(i, j) \in \{0, 1\}^2$, so that we finally obtain:

$$\lim_{n \rightarrow \infty} \frac{\mathbf{B}_n}{\pi^n} = W \frac{\pi}{\pi - 1} \begin{pmatrix} az^0 p_{00} + cz^1 p_{10} \\ az^0 p_{01} + cz^1 p_{11} \end{pmatrix} = W \frac{\pi}{\pi - 1} \mathbf{P}^t \begin{pmatrix} az^0 \\ cz^1 \end{pmatrix} \quad \text{a.s.}$$

and we use Lemma 5.2 to conclude. □

Remark 6.4. Putting together Proposition 6.3 and Eq. (6.5) above, we immediately get that under the same assumptions as that of Proposition 6.3,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} X_k \delta_{2k+i} \delta_{2(2k+i)+j} = \frac{\pi}{\pi - 1} h^i p_{ij} W \quad \text{a.s.}$$

for all $(i, j) \in \{0, 1\}^2$, result we will use for the study of the limit of $\sum X_k^2 \delta_{2k+i}$.

6.3. Asymptotic behavior of the sum of squared observed data

We now turn to the asymptotic behavior of the sums of the squared observed data. Set $K_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} X_k^2$, for i in $\{0, 1\}$, and $\mathbf{K}_n = (K_n^0, K_n^1)^t$. The following result gives the asymptotic behavior of (\mathbf{K}_n) .

Proposition 6.5. *Under assumptions (HN.1), (HN.2), (HI) and (HO), we have the convergence:*

$$\lim_{n \rightarrow \infty} \frac{\mathbf{K}_n}{\pi^n} = \frac{\pi}{\pi - 1} W \mathbf{k} \quad \text{a.s.},$$

where

$$\mathbf{k} = \begin{pmatrix} k^0 \\ k^1 \end{pmatrix} = (\mathbf{I}_2 - \tilde{\mathbf{P}}_2)^{-1} \mathbf{P}^t \begin{pmatrix} (a^2 + \sigma^2)z^0 + \frac{2}{\pi}abh^0 \\ (c^2 + \sigma^2)z^1 + \frac{2}{\pi}cdh^1 \end{pmatrix},$$

and

$$\tilde{\mathbf{P}}_2 = \frac{1}{\pi} \mathbf{P}^t \begin{pmatrix} b^2 & 0 \\ 0 & d^2 \end{pmatrix}.$$

Proof. We use again Equation (2.1) to prove a recursive property for the sequence (\mathbf{K}_n) . Following the same lines as in the proof of Proposition 6.3, we obtain:

$$\frac{\mathbf{K}_n}{\pi^n} = \tilde{\mathbf{P}}_2 \frac{\mathbf{K}_{n-1}}{\pi^{n-1}} + \frac{\mathbf{C}_n}{\pi^n} = \tilde{\mathbf{P}}_2^n \mathbf{K}_0 + \sum_{\ell=1}^n \tilde{\mathbf{P}}_2^{n-\ell} \frac{\mathbf{C}_\ell}{\pi^\ell},$$

where $\mathbf{C}_n = (C_n^0, C_n^1)^t$ is defined by

$$\begin{aligned} C_n^i &= X_1^2 \delta_{2+i} + a^2 \sum_{k \in \mathbb{T}_n^0} \delta_{2k+i} + b^2 \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k} (\delta_{4k+i} - p_{0i}) \\ &+ 2ab \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k} \delta_{4k+i} + 2a \sum_{k \in \mathbb{T}_n^0} \varepsilon_k \delta_{2k+i} + 2b \sum_{k \in \mathbb{T}_n^0} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+i} \\ &+ \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+i} + c^2 \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \delta_{2k+i} + d^2 \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+1} (\delta_{4k+2+i} - p_{1i}) \\ &+ 2cd \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+1} \delta_{4k+2+i} + 2c \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} \varepsilon_k \delta_{2k+i} + 2d \sum_{k \in \mathbb{T}_n^1 \setminus \mathbb{T}_0} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+i}, \end{aligned}$$

for $i \in \{0, 1\}$. Note that $\|\tilde{\mathbf{P}}_2^n\| \leq \pi^{-n} \beta^{2n} \|\mathbf{P}^n\|$, so that $\tilde{\mathbf{P}}_2^n$ converges to 0. In addition, $\sum \|\tilde{\mathbf{P}}_2^n\|$ is bounded, $\mathbf{I}_2 - \tilde{\mathbf{P}}_2$ is invertible and $\sum_{n \geq 0} \tilde{\mathbf{P}}_2^n$ converges to $(\mathbf{I}_2 - \tilde{\mathbf{P}}_2)^{-1}$. In order to use Lemma 5.2, we have to compute the limit of \mathbf{C}_n/π^n . Following the proof of (6.4), we already have, for $(i, j) \in \{0, 1\}^2$,

$$\sum_{k \in \mathbb{T}_n^j} \varepsilon_k \delta_{2k+i} = o(\pi^n) \quad \text{a.s.}$$

We now turn to the terms $\sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij})$, for $(i, j) \in \{0, 1\}^2$. To deal with these terms, we use Lemma 5.1 with the same setting we used to prove Eq. (6.5), except that we replace X_k with X_k^2 . Assumptions (i) and (ii) of Lemma 5.1 have thus already been checked, and regarding (iii), we have $\sum_{k \in \mathbb{T}_{n-1}} X_k^4 \delta_{2k+i} = \mathcal{O}(\pi^n)$ a.s. thanks to Lemma 6.2. We conclude that

$$\sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} (\delta_{2(2k+i)+j} - p_{ij}) = o(\pi^n) \quad \text{a.s.}$$

Next, we study $\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+j}$, for $(i, j) \in \{0, 1\}^2$. We use the same martingale tool, so to speak Lemma 5.1, with $\mathcal{G} = \mathbb{F}^\mathcal{O}$, $H_k = X_{[\frac{k}{2}]} \delta_{2k+j} \mathbb{1}_{\{k \in \mathbb{T}^i\}}$ and $G_k = \varepsilon_k$. Assumptions (i) and (ii) are easily checked, and since

$$\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]}^2 \delta_{2k+j} = \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2(2k+i)+j} \leq \sum_{k \in \mathbb{T}_{n-1}} X_k^2 \delta_{2k+i} = \mathcal{O}(\pi^n),$$

the last equality coming from Lemma 6.1, assumption (iii) is satisfied and

$$\sum_{k \in \mathbb{T}_n^i} X_{[\frac{k}{2}]} \varepsilon_k \delta_{2k+j} = o(\pi^n) \quad \text{a.s.}$$

Now, Corollary 5.6 yields that for $i \in \{0, 1\}$,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n \setminus \mathbb{T}_0} \varepsilon_k^2 \delta_{2k+i} = \sigma^2 (p_{0i} z^0 + p_{1i} z^1) \frac{\pi}{\pi - 1} W \quad \text{a.s.}$$

Finally, Remark 6.4 gives the limit of $\pi^{-n} \sum_{k \in \mathbb{T}_{n-1}} X_k \delta_{2k+i} \delta_{2(2k+i)+j}$, and Lemma 5.3 that of $\pi^{-n} \sum_{k \in \mathbb{T}_n^j} \delta_{2k+i}$, so that we finally obtain

$$\lim_{n \rightarrow \infty} \frac{C_n}{\pi^n} = \frac{W\pi}{\pi - 1} \begin{pmatrix} p_{00} & p_{10} \\ p_{01} & p_{11} \end{pmatrix} \times \begin{pmatrix} (a^2 + \sigma^2)z^0 + \frac{2}{\pi}abh^0 \\ (c^2 + \sigma^2)z^1 + \frac{2}{\pi}cdh^1 \end{pmatrix} \text{ a.s.}$$

And we conclude using Lemma 5.2 again. □

Propositions 6.3 and 6.5 together with Equation (2.7) give the asymptotic behavior of the matrices S_n^0 and S_n^1 . The next result gives the behavior of matrix $S_n^{0,1}$ given through the quantities $\sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k$ and $\sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2$. It is an easy consequence of Propositions 6.3 and 6.5, together with Lemma 5.3 for the first limit.

6.4. Asymptotic behavior of covariance terms

Finally, we turn to the asymptotic behavior of the covariance terms, which are involved in matrix $S_n^{0,1}$. We thus define $H_n^{01} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k$ and $K_n^{01} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} X_k^2$.

Proposition 6.6. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have the almost sure convergences:*

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} = \frac{\pi}{\pi - 1} W \bar{p}(1, 1),$$

$$\lim_{n \rightarrow \infty} \frac{H_n^{01}}{\pi^n} = \frac{\pi}{\pi - 1} W h^{0,1} \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{K_n^{01}}{\pi^n} = \frac{\pi}{\pi - 1} W k^{0,1},$$

where

$$\begin{aligned} \bar{p}(1, 1) &= p^{(0)}(1, 1)z^0 + p^{(1)}(1, 1)z^1, \\ h^{0,1} &= p^{(0)}(1, 1) \left(az^0 + b \frac{h^0}{\pi} \right) + p^{(1)}(1, 1) \left(cz^1 + d \frac{h^1}{\pi} \right), \\ k^{0,1} &= p^{(0)}(1, 1) \left(a^2 z^0 + b^2 \frac{k^0}{\pi} + 2ab \frac{h^0}{\pi} \right) \\ &\quad + p^{(1)}(1, 1) \left(c^2 z^1 + d^2 \frac{k^1}{\pi} + 2cd \frac{h^1}{\pi} \right) + \sigma^2 \bar{p}(1, 1). \end{aligned} \tag{6.6}$$

Proof. The first limit is a consequence of Lemma 5.3. Next, using Eq. (2.1) we obtain $H_n^{01} \pi^{-n}$ and $K_n^{01} \pi^{-n}$ in terms of $\pi^{-n} \sum_{k \in \mathbb{T}_{n-1}^i} \delta_k$, $H_{n-1}^i \pi^{-n}$ and $K_{n-1}^i \pi^{-n}$ and the result follows from Propositions 6.3 and 6.5. □

Proof of Proposition 4.2. We are now in a position to complete the proof of Proposition 4.2. Simply notice that we have proved in Propositions 6.3, 6.5 and 6.6 all the wished convergences, except that we normalized the sums with π^n . Thanks to Lemma 2.1, we end the proof. \square

Remark 6.7. In the case of fully observed data, the matrix \mathbf{P} is a 2×2 matrix with all entries equal to 1, π equals 2 and the normalized eigenvector \mathbf{z} equals $(1/2, 1/2)$. One can check that in that case, our limits correspond to those of [3].

7. Asymptotic behavior of the main martingale

Theorem 4.4 is a strong law of large numbers for the martingale (\mathbf{M}_n) . The standard strong law for martingales is unhelpful here. Indeed, it is valid for martingales that can be decomposed in a sum of the form $\sum_{\ell=1}^n \Psi_{\ell-1} \xi_{\ell}$ where (Ψ_{ℓ}) is predictable and (ξ_{ℓ}) is a martingale difference sequence. In addition, (Ψ_{ℓ}) and (ξ_{ℓ}) are required to be sequences of *fixed-size* vectors. Such a decomposition with fixed-sized vectors is impossible in our context (see Lemma A.2), essentially because the number of observed data in each generation asymptotically grows exponentially fast as π^n . Consequently, we are led to propose a new strong law of large numbers for (\mathbf{M}_n) , adapted to our framework.

For all $n \geq 1$, let $\mathcal{V}_n = \mathbf{M}_n^t \Sigma_{n-1}^{-1} \mathbf{M}_n$ where Σ_n is defined in Section 3.1. First of all, we have

$$\begin{aligned} \mathcal{V}_{n+1} &= (\mathbf{M}_n + \Delta \mathbf{M}_{n+1})^t \Sigma_n^{-1} (\mathbf{M}_n + \Delta \mathbf{M}_{n+1}), \\ &= \mathcal{V}_n - \mathbf{M}_n^t (\Sigma_{n-1}^{-1} - \Sigma_n^{-1}) \mathbf{M}_n + 2 \mathbf{M}_n^t \Sigma_n^{-1} \Delta \mathbf{M}_{n+1} + \Delta \mathbf{M}_{n+1}^t \Sigma_n^{-1} \Delta \mathbf{M}_{n+1}. \end{aligned}$$

Note that $\mathbf{M}_n^t \Sigma_n^{-1} \Delta \mathbf{M}_{n+1}$ and $\Delta \mathbf{M}_{n+1}^t \Sigma_n^{-1} \mathbf{M}_{n+1}$ are scalars, hence they are equal to their own transpose and as a result, one has $\mathbf{M}_n^t \Sigma_n^{-1} \Delta \mathbf{M}_{n+1} = \Delta \mathbf{M}_{n+1}^t \Sigma_n^{-1} \mathbf{M}_{n+1}$. By summing over the identity above, we obtain the main decomposition

$$\mathcal{V}_{n+1} + \mathcal{A}_n = \mathcal{V}_1 + \mathcal{B}_{n+1} + \mathcal{W}_{n+1}, \tag{7.1}$$

where

$$\begin{aligned} \mathcal{A}_n &= \sum_{\ell=1}^n \mathbf{M}_{\ell}^t (\Sigma_{\ell-1}^{-1} - \Sigma_{\ell}^{-1}) \mathbf{M}_{\ell}, \\ \mathcal{B}_{n+1} &= 2 \sum_{\ell=1}^n \mathbf{M}_{\ell}^t \Sigma_{\ell}^{-1} \Delta \mathbf{M}_{\ell+1}, \quad \mathcal{W}_{n+1} = \sum_{\ell=1}^n \Delta \mathbf{M}_{\ell+1}^t \Sigma_{\ell}^{-1} \Delta \mathbf{M}_{\ell+1}. \end{aligned}$$

The asymptotic behavior of the left-hand side of (7.1) is as follows.

Proposition 7.1. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathcal{V}_{n+1} + \mathcal{A}_n}{n} = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Proof. Thanks to the laws of large numbers derived in Sections 5 and 6, the proof of Proposition 7.1 follows essentially the same lines as [3] and is given in Appendix A. \square

Since (\mathcal{V}_n) and (\mathcal{A}_n) are two sequences of non negative real numbers, Proposition 7.1 yields that $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\mathcal{V}_n = \mathcal{O}(n)$ a.s. which proves Equation (4.2). We now turn to the proof of Equation (4.3). We start with a sharp rate of convergence for (\mathbf{M}_n) .

Proposition 7.2. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we, we have, for all $\eta > 1/2$,*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \|\mathbf{M}_n\|^2 = o(|\mathbb{T}_{n-1}^*|n^\eta) \quad a.s.$$

Proof. The result is obvious on \mathcal{E} . On $\bar{\mathcal{E}}$, the proof follows again the same lines as [3] thanks to the laws of large numbers derived in Sections 5 and 6. It is given in Appendix B. \square

A direct application of Proposition 7.2 ensures that $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\mathcal{V}_n = o(n^\eta)$ a.s. for all $\eta > 1/2$. Hence, Proposition 7.1 immediately leads to the following result.

Corollary 7.3. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \frac{\mathcal{A}_n}{n} = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Proof of Result (4.3) of Theorem 4.4. First of all, \mathcal{A}_n may be rewritten as

$$\mathcal{A}_n = \sum_{\ell=1}^n \mathbf{M}_\ell^t (\boldsymbol{\Sigma}_{\ell-1}^{-1} - \boldsymbol{\Sigma}_\ell^{-1}) \mathbf{M}_\ell = \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \boldsymbol{\Delta}_\ell \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \mathbf{M}_\ell,$$

where $\boldsymbol{\Delta}_n = \mathbf{I}_4 - \boldsymbol{\Sigma}_{n-1}^{1/2} \boldsymbol{\Sigma}_n^{-1} \boldsymbol{\Sigma}_{n-1}^{1/2}$. Thanks to Corollary 4.3, we know that

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \boldsymbol{\Delta}_n = \frac{\pi - 1}{\pi} \mathbf{I}_4 \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Besides, Corollary 7.3 yields that $\mathcal{A}_n \sim n \frac{\pi-1}{\pi} 4\sigma^2$ a.s. on $\bar{\mathcal{E}}$. Plugging these two results into the equality

$$\mathcal{A}_n = \frac{\pi - 1}{\pi} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell + \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \left(\boldsymbol{\Delta}_\ell - \frac{\pi - 1}{\pi} \mathbf{I}_4 \right) \boldsymbol{\Sigma}_{\ell-1}^{-1/2} \mathbf{M}_\ell$$

gives that $\sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell \sim \mathcal{A}_n \frac{\pi}{\pi-1}$ a.s. on $\bar{\mathcal{E}}$ and convergence (4.3) directly follows. \square

8. Proof of the main results

We can now proceed to proving our main results.

8.1. Strong consistency for $\widehat{\boldsymbol{\theta}}_n$

Theorem 3.2 is a direct consequence of Theorem 4.4.

Proof of result (3.3) of Theorem 3.2. Recall that $\mathcal{V}_n = \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n$. It clearly follows from Equation (4.1) that

$$\mathcal{V}_n = (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t \boldsymbol{\Sigma}_{n-1} (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}).$$

Consequently, the asymptotic behavior of $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}$ is related to the one of \mathcal{V}_n . More precisely, we can deduce from Corollary 4.3 and the fact that the eigenvalues of a matrix are continuous functions of its coefficients the following result

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\lambda_{\min}(\boldsymbol{\Sigma}_n)}{|\mathbb{T}_n^*|} = \lambda_{\min}(\boldsymbol{\Sigma}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

where $\lambda_{\min}(\mathbf{A})$ denotes the smallest eigenvalue of matrix \mathbf{A} . Since \mathbf{L} as well as $\boldsymbol{\Sigma}$ is definite positive, one has $\lambda_{\min}(\boldsymbol{\Sigma}) > 0$. Therefore, as

$$\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 \leq \frac{\mathcal{V}_n}{\lambda_{\min}(\boldsymbol{\Sigma}_{n-1})},$$

we use Result (4.2) of Theorem 4.4 to conclude that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = \mathcal{O}\left(\frac{n}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\overline{\mathcal{E}}} = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

which completes the proof of results (3.3). □

We now prove the quadratic strong law (QSL).

Proof of result (3.4) of Theorem 3.2. The QSL is a direct consequence of result (4.3) of Theorem 4.4 together with the fact that $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n$. Indeed, we have

$$\begin{aligned} & \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \boldsymbol{\Sigma}_{\ell-1} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \mathbb{1}_{\{|\mathbb{G}_{\ell-1}^*| > 0\}} \frac{\boldsymbol{\Sigma}_{\ell-1}}{|\mathbb{T}_{\ell-1}^*|} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) \\ &= \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \boldsymbol{\Sigma} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) + o(1) \quad \text{a.s.} \end{aligned}$$

which completes the proof. □

8.2. Strong consistency for the variance estimators

For $n \geq 1$, set

$$\mathbf{V}_k = (\delta_{2k}\varepsilon_{2k}, \delta_{2k+1}\varepsilon_{2k+1})^t, \quad \widehat{\mathbf{V}}_k = (\delta_{2k}\widehat{\varepsilon}_{2k}, \delta_{2k+1}\widehat{\varepsilon}_{2k+1})^t.$$

The almost sure convergence of $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n$ is strongly related to that of $\widehat{\mathbf{V}}_k - \mathbf{V}_k$.

Proof of result (3.5) of Theorem 3.3. Equation(3.5) can be rewritten as

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 = 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

Once again, we are searching for a link between the sum of $\|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|$ and the processes (\mathcal{A}_n) and (\mathcal{V}_n) whose convergence properties were previously investigated. For $i \in \{0, 1\}$ and $n \geq 0$, let

$$\Phi_n^i = \begin{pmatrix} \delta_{2(2^n)+i} & \delta_{2(2^{n+1})+i} & \cdots & \delta_{2(2^{n+1}-1)+i} \\ \delta_{2(2^n)+i}X_{2^n} & \delta_{2(2^{n+1})+i}X_{2^{n+1}} & \cdots & \delta_{2(2^{n+1}-1)+i}X_{2^{n+1}-1} \end{pmatrix}$$

be the collection of $(\delta_{2k+i}, \delta_{2k+i}X_k)^t$, $k \in \mathbb{G}_n$, and set

$$\Psi_n = \begin{pmatrix} \Phi_n^0 & 0 \\ 0 & \Phi_n^1 \end{pmatrix}.$$

Note that Ψ_n is a $4 \times 2^{n+1}$ matrix. For all $n \geq 1$, we thus have, in the matrix form

$$\begin{aligned} \sum_{k \in \mathbb{G}_n} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 &= \sum_{k \in \mathbb{G}_n} \delta_{2k}(\widehat{\varepsilon}_{2k} - \varepsilon_{2k})^2 + \delta_{2k+1}(\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1})^2, \\ &= (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})^t \Psi_n \Psi_n^t (\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}), \\ &= \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \Psi_n \Psi_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n, \\ &= \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1/2} \boldsymbol{\Delta}_n \boldsymbol{\Sigma}_{n-1}^{-1/2} \mathbf{M}_n, \end{aligned}$$

where

$$\boldsymbol{\Delta}_n = \boldsymbol{\Sigma}_{n-1}^{-1/2} \Psi_n \Psi_n^t \boldsymbol{\Sigma}_{n-1}^{-1/2} = \boldsymbol{\Sigma}_{n-1}^{-1/2} (\boldsymbol{\Sigma}_n - \boldsymbol{\Sigma}_{n-1}) \boldsymbol{\Sigma}_{n-1}^{-1/2}.$$

Now, we can deduce from Corollary (4.3) that

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \boldsymbol{\Delta}_n = (\pi - 1) \mathbf{I}_4 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

which implies that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \sum_{k \in \mathbb{G}_n} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 = \mathbf{M}_n^t \boldsymbol{\Sigma}_{n-1}^{-1} \mathbf{M}_n (\pi - 1 + o(1)) \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \quad \text{a.s.}$$

Therefore, we can conclude via convergence (4.3) that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 \\ &= \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n(\pi - 1)} \sum_{\ell=1}^n \mathbf{M}_\ell^t \boldsymbol{\Sigma}_{\ell-1}^{-1} \mathbf{M}_\ell = 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

which completes the proof. □

Proof of result (3.6) of Theorem 3.3. First of all, one has

$$\begin{aligned} \widehat{\sigma}_n^2 - \sigma_n^2 &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}} (\|\widehat{\mathbf{V}}_k\|^2 - \|\mathbf{V}_k\|^2), \\ &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_{n-1}} (\|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 + 2(\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k). \end{aligned}$$

Set

$$P_n = \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k.$$

We clearly have

$$\Delta P_{n+1} = P_{n+1} - P_n = \sum_{k \in \mathbb{G}_n} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{V}_k.$$

One can observe that for all $k \in \mathbb{G}_n$, $\widehat{\mathbf{V}}_k - \mathbf{V}_k$ is $\mathcal{F}_n^{\mathcal{O}}$ -measurable. Consequently, (P_n) is a real martingale transform for the filtration $\mathbb{F}^{\mathcal{O}}$. Hence, we can deduce from the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [6] together with (3.5) that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} P_n = o\left(\sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2\right) = o(n) \quad \text{a.s.}$$

It ensures once again via convergence (3.5) that

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\widehat{\sigma}_n^2 - \sigma_n^2) &= \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2 \\ &= 4(\pi - 1)\sigma^2 \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

which completes the proof of result (3.6). □

Proof of results (3.7) and (3.8) of Theorem 3.3. We now turn to the study of the covariance estimator $\widehat{\rho}_n$. We have

$$\begin{aligned} \widehat{\rho}_n - \rho_n &= \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} \widehat{\varepsilon}_{2k+1} - \varepsilon_{2k} \varepsilon_{2k+1}), \\ &= \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \delta_{2k+1} (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) + \frac{1}{|\mathbb{T}_{n-1}^{*01}|} Q_n, \end{aligned}$$

where

$$\begin{aligned} Q_n &= \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) \varepsilon_{2k+1} + \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \varepsilon_{2k} \\ &= \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\mathbf{V}}_k - \mathbf{V}_k)^t \mathbf{J}_2 \mathbf{V}_k, \end{aligned}$$

with

$$\mathbf{J}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The process (Q_n) is a real martingale transform for the filtration $\mathbb{F}^{\mathcal{O}}$ satisfying

$$Q_n = o\left(\sum_{k \in \mathbb{T}_{n-1}} \|\widehat{\mathbf{V}}_k - \mathbf{V}_k\|^2\right) = o(n) \quad \text{a.s.}$$

It now remains to prove that

$$\begin{aligned} &\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} \delta_{2k+1} (\widehat{\varepsilon}_{2k} - \varepsilon_{2k}) (\widehat{\varepsilon}_{2k+1} - \varepsilon_{2k+1}) \\ &= \lim_{n \rightarrow \infty} \frac{R_n}{n} = \rho(\pi - 1) \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned} \tag{8.1}$$

where

$$R_n = \sum_{\ell=1}^n \mathbf{M}_{\ell}^t \Sigma_{\ell-1}^{-1} (\mathbf{J}_2 \otimes \Phi_{\ell}^{01} (\Phi_{\ell}^{01})^t) \Sigma_{\ell-1}^{-1} \mathbf{M}_{\ell},$$

where \otimes denotes the Kronecker product of matrices, i.e.

$$\mathbf{J}_2 \otimes \Phi_{\ell}^{01} (\Phi_{\ell}^{01})^t = \begin{pmatrix} 0 & \Phi_{\ell}^{01} (\Phi_{\ell}^{01})^t \\ \Phi_{\ell}^{01} (\Phi_{\ell}^{01})^t & 0 \end{pmatrix},$$

and Φ_{ℓ}^{01} is defined similarly as Φ_{ℓ}^0 and Φ_{ℓ}^1 by the collection of $(\delta_{2k} \delta_{2k+1}, \delta_{2k} \delta_{2k+1} X_k)^t$, $k \in \mathbb{G}_{\ell}$. As $\Phi_n^{01} (\Phi_n^{01})^t = \mathbf{S}_n^{01} - \mathbf{S}_{n-1}^{01}$, proposition 4.2 implies that

$$\lim_{n \rightarrow \infty} \Sigma_{n-1}^{-1/2} (\mathbf{J}_2 \otimes \Phi_n^{01} (\Phi_n^{01})^t) \Sigma_{n-1}^{-1/2} = (\pi - 1) \Sigma^{-1/2} \mathbf{J}_2 \otimes \mathbf{L}^{01} \Sigma^{-1/2} \quad \text{a.s.}$$

so that the asymptotic behavior of R_n/n boils down to that of

$$\sum_{\ell=1}^n \mathbf{M}_{\ell}^t \Sigma_{\ell-1}^{-1/2} (\mathbf{J}_2 \otimes \mathbf{L}^{01}) \Sigma_{\ell-1}^{-1/2} \mathbf{M}_{\ell}.$$

A proof along the same lines as in Section 7 finally yields the expected results, i.e.

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{R_n}{n} = \rho \frac{\pi - 1}{\pi} \text{tr}((\mathbf{L}^1)^{-1} (\mathbf{L}^{0,1})^2 (\mathbf{L}^0)^{-1}) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

which completes the proof of convergence (8.1). We then obtain

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{T}_n^*| > 0\}} \frac{|\mathbb{T}_n^*|}{n} (\hat{\rho}_n - \rho_n) = \rho \frac{\pi - 1}{\bar{p}(1, 1)} \text{tr}((\mathbf{L}^1)^{-1}(\mathbf{L}^{0,1})^2(\mathbf{L}^0)^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

which completes the proof of Theorem 3.3. □

8.3. Asymptotic normality

Contrary to the previous literature on BAR processes, we cannot use the central limit theorem given by Propositions 7.8 and 7.9 of [11] as in [8, 3] because the normalizing term is now the number of observations and is therefore random. The approach used in [5] strongly relies on the gaussian assumption for the noise sequence that does not hold here. Instead, we use the central limit theorem for martingales given in Theorem 3.II.10 of Duflo [6]. However, unlike the previous sections, this theorem can not be directly applied to the martingale (M_n) because the number of observed data in a given generation grows exponentially fast and the Lindeberg condition does not hold. The solution is to use a new filtration. Namely, instead of using the observed generation-wise filtration, we will use the sister pair-wise one. Let

$$\mathcal{G}_p^{\mathcal{O}} = \mathcal{O} \vee \sigma\{\delta_1 X_1, (\delta_{2k} X_{2k}, \delta_{2k+1} X_{2k+1}), 1 \leq k \leq p\}$$

be the σ -algebra generated by the whole history \mathcal{O} of the Galton-Watson process and all observed individuals up to the offspring of individual p . Hence $(\delta_{2k} \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1})$ is $\mathcal{G}_k^{\mathcal{O}}$ -measurable. In addition, assumptions **(HN.1)** and **(HI)** imply that the processes $(\delta_{2k} \varepsilon_{2k}, X_k \delta_{2k} \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1}, X_k \delta_{2k+1} \varepsilon_{2k+1})^t$, $(\delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{2k+1}^2 - (\delta_{2k} + \delta_{2k+1}) \sigma^2)$ and $(\delta_{2k} \delta_{2k+1} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho))$ are $\mathcal{G}_k^{\mathcal{O}}$ -martingale difference sequences. In all the sequel, we will work under the probability $\mathbb{P}_{\bar{\mathcal{E}}}$ and we denote by $\mathbb{E}_{\bar{\mathcal{E}}}$ the corresponding expectation.

Proof of Theorem 3.4, first step. We apply Theorem 3.II.10 of [6] to the $\mathcal{G}_k^{\mathcal{O}}$ -martingale $M^{(n)} = (M_p^{(n)})_{\{p \geq 1\}}$ defined by

$$M_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^p D_k \quad \text{with} \quad D_k = \begin{pmatrix} \delta_{2k} \varepsilon_{2k} \\ X_k \delta_{2k} \varepsilon_{2k} \\ \delta_{2k+1} \varepsilon_{2k+1} \\ X_k \delta_{2k+1} \varepsilon_{2k+1} \end{pmatrix}.$$

Set $\nu_n = |\mathbb{T}_n| = 2^{n+1} - 1$. Note that if $k \notin \mathbb{T}_n^*$, then $D_k = 0$ which implies that

$$M_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^{|\mathbb{T}_n|} D_k = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k \in \mathbb{T}_n^*} D_k.$$

As the non-extinction set $\bar{\mathcal{E}}$ is in $\mathcal{G}_k^{\mathcal{O}}$ for every $k \geq 1$, it is easy to prove that

$$\begin{aligned} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] &= \mathbb{E}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] \\ &= \begin{pmatrix} \sigma^2 \delta_{2k} & \sigma^2 \delta_{2k} X_k & \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k \\ \sigma^2 \delta_{2k} X_k & \sigma^2 \delta_{2k} X_k^2 & \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 \\ \rho \delta_{2k} \delta_{2k+1} & \rho \delta_{2k} \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} & \sigma^2 \delta_{2k+1} X_k \\ \rho \delta_{2k} \delta_{2k+1} X_k & \rho \delta_{2k} \delta_{2k+1} X_k^2 & \sigma^2 \delta_{2k+1} X_k & \sigma^2 \delta_{2k+1} X_k^2 \end{pmatrix}, \end{aligned}$$

and Corollary 4.3 gives the $\mathbb{P}_{\bar{\mathcal{E}}}$ almost sure limit of the increasing process

$$\langle M^{(n)} \rangle_{\nu_n} = \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] = \frac{\mathbf{\Gamma}_n}{|\mathbb{T}_n^*|} \xrightarrow{n \rightarrow \infty} \mathbf{\Gamma}. \tag{8.2}$$

Therefore, the first assumption of Theorem 3.II.10 of [6] holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. We now want to prove the Lindeberg condition that is the convergence in probability to 0 of the following expression L_n for all $\epsilon > 0$:

$$\begin{aligned} L_n &= \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|\mathbf{D}_k\|^2 \mathbb{1}_{\{\|\mathbf{D}_k\| > \epsilon \sqrt{|\mathbb{T}_n^*|}\}} | \mathcal{G}_{k-1}^{\mathcal{O}}] \\ &\leq \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|\mathbf{D}_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] \mathbb{P}_{\bar{\mathcal{E}}}(\|\mathbf{D}_k\| > \epsilon \sqrt{|\mathbb{T}_n^*|} | \mathcal{G}_{k-1}^{\mathcal{O}}) \\ &\leq \frac{\sup_{k \geq 0} \mathbb{E}[\|\mathbf{D}_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}]}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \frac{\mathbb{E}_{\bar{\mathcal{E}}}[\|\mathbf{D}_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}]}{\epsilon^2 |\mathbb{T}_n^*|}, \end{aligned}$$

for some $r > 2$ and thanks to Hölder and Chebyshev inequalities. Besides, using Eq. (6.1) and similar calculations as in Lemma 6.1, one readily obtains

$$X_n^8 \leq 2^7 (1 - \beta)^{-7} \sum_{k=0}^{r_n-1} \beta^k |\eta_{\lfloor \frac{n}{2^k} \rfloor}^8| + 2^7 \beta^{8r_n} X_1^8.$$

Now, assumption (HN.1) together with $\beta < 1$ yield the existence of a constant C such that

$$\sup_{k \geq 0} \mathbb{E}[X_k^8] \leq C(1 + \mathbb{E}[X_1^8]),$$

and recall that $\mathbb{E}[X_1^8] < \infty$. Finally, since the entries of \mathbf{D}_k are combinations of ε_{2k+i} and X_k , using again (HN.1) and (HI), one obtains that

$$\sup_{k \geq 0} \mathbb{E}[\|\mathbf{D}_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty \quad \text{a.s.}$$

with $r = 8$. The Lindeberg condition is thus proved, plugging the convergence (8.2) into the following equality:

$$\frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\|\mathbf{D}_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}] = \text{tr} \left(\frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] \right) \xrightarrow{n \rightarrow \infty} \text{tr}(\mathbf{\Gamma}).$$

We can now conclude that under $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} \sum_{k \in \mathbb{T}_{n-1}^*} D_k = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} M_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

Finally, result (3.9) follows from Eq. (4.1) and Corollary 4.3 together with Slutsky's Lemma. \square

Proof of Theorem 3.4, second step. On the one hand, we apply Theorem 3.II.10 of [6] to the $\mathcal{G}_p^{\mathcal{O}}$ -martingale $M^{(n)} = (M_p^{(n)})_{\{p \geq 1\}}$ defined by

$$M_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k=1}^p v_k \quad \text{and} \quad v_k = \delta_{2k} \varepsilon_{2k}^2 + \delta_{2k+1} \varepsilon_{k+1}^2 - (\delta_{2k} + \delta_{2k+1}) \sigma^2.$$

As above, one clearly has

$$M_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_n^*|}} \sum_{k \in \mathbb{T}_{n-1}^*} v_k = \sqrt{|\mathbb{T}_n^*|} (\sigma_n^2 - \sigma^2).$$

Using assumptions (HN.1), (HI) and Lemma 5.3 we compute the limit of the increasing process under $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\lim_{n \rightarrow \infty} \langle M^{(n)} \rangle_{\nu_n} = (\tau^4 - \sigma^4) + \frac{2\bar{p}(1, 1)}{\pi} (\nu^2 \tau^4 - \sigma^4) \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

Therefore, the first assumption of Theorem 3.II.10 of [6] holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. Thanks to assumptions (HN.2) and (HI) we can prove that for some $r > 2$,

$$\sup_{k \geq 0} \mathbb{E}_{\bar{\mathcal{E}}} [\|v_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty \quad \text{a.s.}$$

which implies the Lindeberg condition. Therefore, we obtain that under $\mathbb{P}_{\bar{\mathcal{E}}}$

$$\sqrt{|\mathbb{T}_n^*|} (\sigma_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\tau^4 - \sigma^4) + \frac{2\bar{p}(1, 1)}{\pi} (\nu^2 \tau^4 - \sigma^4)).$$

Furthermore, we infer from Eq. (3.6) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_n^*|} (\hat{\sigma}_n^2 - \sigma_n^2) = 0 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

which yields result (3.10).

We turn now to the proof of result (3.11) with another $\mathcal{G}_p^{\mathcal{O}}$ -martingale $(M^{(n)})$ defined by

$$M_p^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^{*01}|}} \sum_{k=1}^p \delta_{2k} \delta_{2k+1} (\varepsilon_{2k} \varepsilon_{2k+1} - \rho).$$

As above, one easily shows that

$$M_{\nu_n}^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^{*01}|}} \sum_{i \in \mathbb{T}_{n-1}^*} \delta_{2i} \delta_{2i+1} (\varepsilon_{2i} \varepsilon_{2i+1} - \rho) = \sqrt{|\mathbb{T}_{n-1}^{*01}|} (\rho_n - \rho).$$

Using assumptions **(HN.1)** and **(H.I)**, we compute the limit of the increasing process

$$\lim_{n \rightarrow \infty} \langle M^{(n)} \rangle_{\nu_n} = \nu^2 \tau^4 - \rho^2 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

We also derive the Lindeberg condition. Consequently, we obtain that under $\mathbb{P}_{\bar{\mathcal{E}}}$, one has

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|}(\rho_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 \tau^4 - \rho^2).$$

Furthermore, we infer from (3.8) that

$$\lim_{n \rightarrow \infty} \sqrt{|\mathbb{T}_{n-1}^{*01}|}(\hat{\rho}_n - \rho_n) = 0 \quad \mathbb{P}_{\bar{\mathcal{E}}} \text{ a.s.}$$

Finally, result (3.11) follows, which completes the proof of Theorem 3.4. \square

Appendix A: Quadratic strong law

In order to establish the quadratic strong law for the main martingale (M_n) , we are going to study separately the asymptotic behavior of (W_n) and (B_n) which appear in the main decomposition given by Equation (7.1).

Lemma A.1. *Under assumptions **(HN.1)**, **(HN.2)**, **(HO)** and **(HI)**, we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} W_n = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

Proof. First of all, we have the decomposition $W_{n+1} = \mathcal{T}_{n+1} + \mathcal{R}_{n+1}$ where

$$\begin{aligned} \mathcal{T}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t \Sigma^{-1} \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}, \\ \mathcal{R}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t (|\mathbb{T}_{\ell}^*| \Sigma_{\ell}^{-1} - \Sigma^{-1}) \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}. \end{aligned}$$

We first prove that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{T}_n = \frac{4(\pi - 1)}{\pi} \sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \tag{A.1}$$

As \mathcal{T}_n is a scalar and the trace is commutative, one can rewrite \mathcal{T}_{n+1} as $\mathcal{T}_{n+1} = \text{tr}(\Sigma^{-1/2} \mathcal{H}_{n+1} \Sigma^{-1/2})$ where

$$\mathcal{H}_{n+1} = \sum_{\ell=1}^n \frac{\Delta M_{\ell+1} \Delta M_{\ell+1}^t}{|\mathbb{T}_{\ell}^*|}.$$

Our goal is to make use of the strong law of large numbers for martingale transforms, so we start by adding and subtracting a term involving the conditional expectation of $\Delta \mathcal{H}_{n+1}$ given $\mathcal{F}_n^{\mathcal{O}}$. We have already seen in Section 4.1 that for

all n , $\mathbb{E}[\Delta M_{n+1} \Delta M_{n+1}^t | \mathcal{F}_n^{\mathcal{O}}] = \Gamma_n - \Gamma_{n-1}$. Consequently, we can split \mathcal{H}_{n+1} into two terms

$$\mathcal{H}_{n+1} = \sum_{\ell=1}^n \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} + \mathcal{K}_{n+1},$$

where

$$\mathcal{K}_{n+1} = \sum_{\ell=1}^n \frac{\Delta M_{\ell+1} \Delta M_{\ell+1}^t - (\Gamma_{\ell} - \Gamma_{\ell-1})}{|\mathbb{T}_{\ell}^*|}.$$

On the one hand, it follows from Corollary 4.3 and Lemma 2.1 that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Gamma_n - \Gamma_{n-1}}{|\mathbb{T}_n^*|} = \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

Thus, Cesaro convergence and the remark that $\{|\mathbb{G}_{\ell}^*| = 0\} \subset \{|\mathbb{G}_n^*| = 0\}$ for all $\ell \leq n$ yield

$$\begin{aligned} \lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} &= \lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_{\ell}^*| > 0\}} \frac{\Gamma_{\ell} - \Gamma_{\ell-1}}{|\mathbb{T}_{\ell}^*|} \\ &= \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

On the other hand, the sequence (\mathcal{K}_n) is obviously a matrix martingale transform and tedious but straightforward calculations, together with Lemmas 6.1 and 6.2 and the strong law of large numbers for martingale transforms given in Theorem 1.3.24 of [6] imply that $\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathcal{K}_n = o(n)$ a.s. Hence, we infer from the equation above that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{H}_n = \frac{\pi - 1}{\pi} \Gamma \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \tag{A.2}$$

Finally, we obtain

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathcal{T}_n = \frac{\pi - 1}{\pi} \text{tr}(\Sigma^{-1/2} \Gamma \Sigma^{-1/2}) \mathbb{1}_{\bar{\mathcal{E}}} = \frac{\pi - 1}{\pi} 4\sigma^2 \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

which proves (A.1). We now turn to the asymptotic behavior of \mathcal{R}_{n+1} . We know from Proposition 4.2 that $\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} (|\mathbb{T}_n^*| \Sigma_n^{-1} - \Sigma^{-1})$ goes to 0 as n goes to infinity. Hence, for all positive ϵ and for large enough n , one has

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\Delta M_{n+1}^t (|\mathbb{T}_n^*| \Sigma_n^{-1} - \Sigma^{-1}) \Delta M_{n+1}| \leq \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} 4\epsilon \Delta M_{n+1}^t \Delta M_{n+1}.$$

Using again that $\{|\mathbb{G}_{\ell}^*| = 0\} \subset \{|\mathbb{G}_{n+1}^*| = 0\}$ for all $\ell \leq n + 1$, we rewrite

$$\mathbb{1}_{\{|\mathbb{G}_{n+1}^*| > 0\}} \mathcal{R}_{n+1} = \mathbb{1}_{\{|\mathbb{G}_{n+1}^*| > 0\}} \sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_{\ell}^*| > 0\}} \frac{\Delta M_{\ell+1}^t (|\mathbb{T}_{\ell}^*| \Sigma_{\ell}^{-1} - \Sigma^{-1}) \Delta M_{\ell+1}}{|\mathbb{T}_{\ell}^*|}.$$

Hence,

$$\begin{aligned} \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}}|\mathcal{R}_{n+1}| &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}}\sum_{\ell=1}^n \mathbb{1}_{\{|\mathbb{G}_\ell^*|>0\}}\frac{\Delta M_{\ell+1}^t \Delta M_{\ell+1}}{|\mathbb{T}_\ell^*|} \\ &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}}\sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t \Delta M_{\ell+1}}{|\mathbb{T}_\ell^*|} \\ &\leq 4\epsilon \mathbb{1}_{\{|\mathbb{G}_{n+1}^*|>0\}}tr(\mathcal{H}_{n+1}). \end{aligned}$$

This last inequality holding for any positive ϵ and large enough n , the limit given by Equation (A.2) entails that

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\frac{1}{n}\mathcal{R}_n = 0 \quad \text{a.s.}$$

which completes the proof of Lemma A.1. □

Lemma A.2. *Under assumptions (HN.1), (HN.2), (HO) and (HI), we have*

$$\lim_{n \rightarrow +\infty} \mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}\frac{1}{n}\mathcal{B}_n = 0 \quad \text{a.s.}$$

Proof. The result is obvious on the extinction set \mathcal{E} . Now let us work on $\bar{\mathcal{E}}$. Now for $i \in \{0, 1\}$ and $n \geq 1$, let $\xi_n^i = (\varepsilon_{2n+i}, \varepsilon_{2n+2+i}, \dots, \varepsilon_{2n+1-2+i})^t$, be the collection of ε_k , $k \in \mathbb{G}_n^i$, and set $\xi_n = (\xi_n^0, \xi_n^1)^t$. Note that ξ_n is a column vector of size 2^{n+1} . With these notation, one has

$$\mathcal{B}_{n+1} = 2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Delta M_{\ell+1} = 2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell \xi_{\ell+1}.$$

The sequence (\mathcal{B}_n) is a real martingale transform satisfying

$$\Delta \mathcal{B}_{n+1} = \mathcal{B}_{n+1} - \mathcal{B}_n = 2M_n^t \Sigma_n^{-1} \Psi_n \xi_{n+1}.$$

Consequently, via the strong law of large numbers for martingale transforms, we find that either (\mathcal{B}_n) converges a.s. or $\mathcal{B}_n = o(\langle \mathcal{B} \rangle_n)$ a.s. where

$$\langle \mathcal{B} \rangle_{n+1} = 4 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell C \Psi_\ell^t \Sigma_\ell^{-1} M_\ell,$$

with

$$C = \begin{pmatrix} \sigma^2 & \rho \\ \rho & \sigma^2 \end{pmatrix} \otimes \mathbf{I}_{2^n}.$$

As C is definite positive under assumption (HN.1), one has $C \leq 2\sigma^2 \mathbf{I}_{2^{n+1}}$ in the sense that $2\sigma^2 \mathbf{I}_{2^{n+1}} - C$ is semi definite positive. Hence, one has

$$\langle \mathcal{B} \rangle_{n+1} \leq 8\sigma^2 \sum_{\ell=1}^n M_\ell^t \Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} M_\ell.$$

Now, by definition, one has

$$\Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} = \begin{pmatrix} (\mathbf{S}_\ell^0)^{-1} \Phi_\ell^0 (\Phi_\ell^0)^t (\mathbf{S}_\ell^0)^{-1} & 0 \\ 0 & (\mathbf{S}_\ell^1)^{-1} \Phi_\ell^1 (\Phi_\ell^1)^t (\mathbf{S}_\ell^1)^{-1} \end{pmatrix}.$$

We now use Lemma B.1 of [3] on each entry to obtain

$$\Sigma_\ell^{-1} \Psi_\ell \Psi_\ell^t \Sigma_\ell^{-1} \leq \Sigma_{\ell-1}^{-1} - \Sigma_\ell^{-1},$$

as the matrix l_k in that lemma is definite positive. Therefore, we obtain that

$$\langle \mathcal{B} \rangle_{n+1} \leq 8\sigma^2 \sum_{\ell=1}^n \mathbf{M}_\ell^t (\Sigma_{\ell-1}^{-1} - \Sigma_\ell^{-1}) \mathbf{M}_\ell = 8\sigma^2 \mathcal{A}_n.$$

Finally, we deduce from the main decomposition given by Equation (7.1) and Lemma A.1 that

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} (\mathcal{V}_{n+1} + \mathcal{A}_n) = o(\mathcal{A}_n) + \mathcal{O}(n) \quad \text{a.s.}$$

leading to $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{V}_{n+1} = \mathcal{O}(n)$ and $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{A}_n = \mathcal{O}(n)$ a.s. as \mathcal{V}_{n+1} and \mathcal{A}_n are non-negative. This implies in turn that $\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}} \mathcal{B}_n = o(n)$ a.s. completing the proof of Lemma A.2. \square

Appendix B: Wei’s Lemma

In order to prove Proposition 7.2, we shall apply Wei’s Lemma given in [16] page 1672, to each entry of the vector-valued main martingale

$$\mathbf{M}_n = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} (\delta_{2k} \varepsilon_{2k}, \delta_{2k} X_k \varepsilon_{2k}, \delta_{2k+1} \varepsilon_{2k+1}, \delta_{2k+1} X_k \varepsilon_{2k+1})^t.$$

For $i \in \{0, 1\}$, denote

$$P_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k+i} \varepsilon_{2k+i} \quad \text{and} \quad Q_n^i = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{2k+i} X_k \varepsilon_{2k+i}.$$

On the set $\bar{\mathcal{E}}$, these processes can be rewritten as

$$P_n^i = \sum_{\ell=1}^n \sqrt{|\mathbb{G}_{\ell-1}^*|} v_\ell^i, \quad Q_n^i = \sum_{\ell=1}^n \sqrt{|\mathbb{G}_{\ell-1}|} w_\ell^i,$$

where

$$v_n^i = \mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}} \frac{1}{\sqrt{|\mathbb{G}_{n-1}^*|}} \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i} \varepsilon_{2k+i},$$

$$w_n^i = \mathbb{1}_{\{|\mathbb{G}_{n-1}^*|>0\}} \frac{1}{\sqrt{|\mathbb{G}_{n-1}^*|}} \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i} X_k \varepsilon_{2k+i}.$$

On the one hand, we clearly have $\mathbb{E}[v_{n+1}^i | \mathcal{F}_n^{\mathcal{O}}] = 0$ and $\mathbb{E}[(v_{n+1}^i)^2 | \mathcal{F}_n^{\mathcal{O}}] = \sigma^2 \frac{Z_{n+1}^i}{|\mathbb{G}_n^*|}$ a.s. on $\bar{\mathcal{E}}$. Moreover, it follows from Cauchy-Schwarz inequality that

$$\begin{aligned} \mathbb{E}[(v_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] &= \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \sum_{k \in \mathbb{G}_n} \delta_{2l+i} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \\ &\quad + \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \sum_{p \in \mathbb{G}_n} \sum_{k \neq p} \delta_{2p+i} \delta_{2k+i} \mathbb{E}[\varepsilon_{2p+i}^2 | \mathcal{F}_n^{\mathcal{O}}] \mathbb{E}[\varepsilon_{2k+i}^2 | \mathcal{F}_n^{\mathcal{O}}] \\ &\leq 3C \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \sup_{k \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \quad \text{a.s.} \end{aligned}$$

as $Z_{n+1}^i |\mathbb{G}_n^*|^{-1}$ is bounded. This implies that $\sup \mathbb{E}[(v_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] < +\infty$ a.s. Consequently, we deduce from Wei's Lemma that for all $\eta > 1/2$,

$$\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}} (P_n^i)^2 = o(|\mathbb{T}_{n-1}^*| n^\eta) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

On the other hand, it is not hard to see that $\mathbb{E}[w_{n+1}^i | \mathcal{F}_n^{\mathcal{O}}] = 0$ a.s. Moreover, it follows from Cauchy-Schwarz inequality that,

$$\begin{aligned} \mathbb{E}[(w_{n+1}^i)^4 | \mathcal{F}_n^{\mathcal{O}}] &\leq \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{G}_n^*|^2} \left(\sum_{k \in \mathbb{G}_n} \delta_{2k+i} X_k^4 \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] + \sigma^4 \sum_{p \in \mathbb{G}_n} \sum_{k \neq p} \delta_{2p+i} \delta_{2k+i} X_p^2 X_k^2 \right) \\ &\leq 3 \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \left(\sup_{k \in \mathbb{G}_n} \mathbb{E}[\varepsilon_{2k+i}^4 | \mathcal{F}_n^{\mathcal{O}}] \right) \left(\frac{1}{|\mathbb{G}_n^*|} \sum_{k \in \mathbb{G}_n} \delta_{2k+i} X_l^2 \right)^2 \quad \text{a.s.} \end{aligned}$$

which is finite from Proposition 6.5. We deduce from Wei's Lemma applied to Q_n^i that for all $\eta > 1/2$, $\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}} \|Q_n^i\|^2 = o(|\mathbb{T}_{n-1}^*| n^\eta)$ a.s. which completes the proof of Proposition 7.2. \square

References

- [1] KRISHNA B. ATHREYA AND PETER E. NEY. *Branching processes*. Springer-Verlag, New York, 1972. Die Grundlehren der mathematischen Wissenschaften, Band 196. [MR0373040](#)
- [2] VINCENT BANSAYE. Proliferating parasites in dividing cells: Kimmel's branching model revisited. *Ann. Appl. Probab.*, 18(3):967–996, 2008. [MR2418235](#)
- [3] BERNARD BERCU, BENOÎTE DE SAPORTA, AND ANNE GÉGOUT-PETIT. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.*, 14:no. 87, 2492–2526, 2009. [MR2563249](#)
- [4] R. COWAN AND R. G. STAUDTE. The bifurcating autoregressive model in cell lineage studies. *Biometrics*, 42:769–783, 1986.
- [5] JEAN-FRANÇOIS DELMAS AND LAURENCE MARSALLE. Detection of cellular aging in a Galton-Watson process. *Stoch. Process. and Appl.*, 120:2495–2519, 2010. [MR2728175](#)

- [6] MARIE DUFLO. *Random iterative models*, volume 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997. [MR1485774](#)
- [7] STEVEN N. EVANS AND DAVID STEINSALTZ. Damage segregation at fissioning may increase growth rates: A superprocess model. *Theoretical Population Biology*, 71(4):473–490, 2007.
- [8] JULIEN GUYON. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.*, 17(5-6):1538–1569, 2007. [MR2358633](#)
- [9] JULIEN GUYON, ARIANE BIZE, GRÉGORIE PAUL, ERIC STEWART, JEAN-FRANÇOIS DELMAS, AND FRANÇOIS TADDÉI. Statistical study of cellular aging. In *CEMRACS 2004—mathematics and applications to biology and medicine*, volume 14 of *ESAIM Proc.*, pages 100–114 (electronic). EDP Sci., Les Ulis, 2005. [MR2226805](#)
- [10] PATSY HACCOU, PETER JAGERS, AND VLADIMIR A. VATUTIN. *Branching processes: variation, growth, and extinction of populations*. Cambridge Studies in Adaptive Dynamics. Cambridge University Press, Cambridge, 2007. [MR2429372](#)
- [11] JAMES D. HAMILTON. *Time series analysis*. Princeton University Press, Princeton, NJ, 1994. [MR1278033](#)
- [12] THEODORE E. HARRIS. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963. [MR0163361](#)
- [13] ROGER A. HORN AND CHARLES R. JOHNSON. *Matrix analysis*. Cambridge University Press, Cambridge, 1990. Corrected reprint of the 1985 original. [MR1084815](#)
- [14] F. MAAOUIA AND A. TOUATI. Identification of multitype branching processes. *Ann. Statist.*, 33(6):2655–2694, 2005. [MR2253098](#)
- [15] E.J. STEWART, R. MADDEN, G. PAUL, AND F. TADDEI. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.*, 3(2):e45, 2005.
- [16] C. Z. WEI. Adaptive prediction by least squares predictors in stochastic regression models with applications to time series. *Ann. Statist.*, 15(4):1667–1682, 1987. [MR0913581](#)

- 1.3 [dSGP⁺12] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Asymmetry tests for bifurcating autoregressive processes with missing data. *Statistics & Probability Letters*, 82(7) :1439–1444, 2012.



ELSEVIER

Contents lists available at SciVerse ScienceDirect

Statistics and Probability Letters

journal homepage: www.elsevier.com/locate/stapro

Asymmetry tests for bifurcating auto-regressive processes with missing data

Benoîte de Saporta^{a,*}, Anne Gégout-Petit^b, Laurence Marsalle^c

^a Univ. Bordeaux, GREThA CNRS UMR 5113, IMB CNRS UMR 5251, INRIA Bordeaux Sud Ouest team CQFD, 351 cours de la Libération, 33405 Talence Cedex, France

^b Univ. Bordeaux, IMB CNRS UMR 5251, INRIA Bordeaux Sud Ouest team CQFD, 351 cours de la Libération, 33405 Talence Cedex, France

^c Univ. Lille 1, Laboratoire Paul Painlevé, CNRS UMR 8524, 59655 Villeneuve d'Ascq Cedex, France

ARTICLE INFO

Article history:

Received 16 December 2011
Received in revised form 4 April 2012
Accepted 4 April 2012
Available online 11 April 2012

MSC:

62M07
62F05
60J80
62P10

Keywords:

Bifurcating autoregressive processes
Cell division data
Missing data
Two-type Galton–Watson model
Wald's test

ABSTRACT

We present symmetry tests for bifurcating autoregressive (BAR) processes when some data are missing. BAR processes typically model cell division data. Each cell can be of one of two types *odd* or *even*. The goal of this paper is to study the possible asymmetry between odd and even cells in a single observed lineage. We first derive asymmetry tests for the lineage itself, modeled by a two-type Galton–Watson process, and then derive tests for the observed BAR process. We present applications on simulated and real data.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Bifurcating autoregressive (BAR) processes were first introduced by Cowan and Staudte (1986). They generalize autoregressive processes when data are structured as a binary tree, see also Hwang and Basawa (2009, 2011) for processes indexed by general trees. Typically, BAR processes are involved in statistical studies of cell lineages. Cell lineage data consist of observations of some quantitative characteristic of the cells over several generations descended from an initial cell. One may need to distinguish the two offspring of a given cell according to some biological property, leading to the notion of *type*. The initial cell is labeled 1, and the two offspring of cell k are labeled $2k$ and $2k + 1$, where $2k$ is of type *even*, and $2k + 1$ is of type *odd*. If X_k denotes the quantitative characteristic of cell k , the first-order asymmetric BAR process is given by

$$\begin{cases} X_{2k} = a + bX_k + \varepsilon_{2k}, \\ X_{2k+1} = c + dX_k + \varepsilon_{2k+1}, \end{cases} \quad (1)$$

for all $k \geq 1$. The noise sequence $(\varepsilon_{2k}, \varepsilon_{2k+1})$ represents environmental effects, while a, b, c, d are unknown real parameters related to the inherited effects. Various estimators are studied in the literature for a, b, c, d , see Guyon (2007), Delmas and

* Corresponding author. Tel.: +33 5 24 57 41 69; fax: +33 5 24 57 40 24.

E-mail addresses: saporta@math.u-bordeaux1.fr (B. de Saporta), anne.petit@u-bordeaux2.fr (A. Gégout-Petit), laurence.marsalle@univ-lille1.fr (L. Marsalle).

Marsalle (2010), Bercu et al. (2009) and de Saporta et al. (2011). Here the genealogy is modeled by a two-type Galton–Watson process (GW), allowing the reproduction laws to depend on both the mother’s and daughter’s types. The aim of this paper is to propose asymmetry tests for both the GW process defining the genealogy of the cells, and the BAR process with missing data. In particular, we propose original tailor-made estimators for the reproduction probabilities of the GW process, give their asymptotic behavior and derive Wald test for the equality of the means of the reproduction laws for even and odd mother cells. We also investigate asymmetry of the parameters a, b, c, d . A detailed study on simulated data, as well as a new investigation of the *Escherichia coli* data of Stewart et al. (2005) are provided.

2. Notation

For all $n \geq 1$, denote the n -th generation by $\mathbb{G}_n = \{k \in \mathbb{N} : 2^n \leq k \leq 2^{n+1} - 1\}$, and the sub-tree of all cells up to the n -th generation by $\mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$. The cardinality $|\mathbb{G}_n|$ of \mathbb{G}_n is 2^n , while that of \mathbb{T}_n is $|\mathbb{T}_n| = 2^{n+1} - 1$. We encode the presence or absence of cells by the process (δ_k) : if cell k is observed, $\delta_k = 1$, if cell k is missing, $\delta_k = 0$. We define the sets of observed cells as $\mathbb{G}_n^* = \{k \in \mathbb{G}_n : \delta_k = 1\}$ and $\mathbb{T}_n^* = \{k \in \mathbb{T}_n : \delta_k = 1\}$. Finally, let \mathcal{E} be the event corresponding to the case when there are no cell left to observe: $\mathcal{E} = \bigcup_{n \geq 1} \{|\mathbb{G}_n^*| = 0\}$ and $\bar{\mathcal{E}}$ its complementary set. For $n \geq 1$, we define the number of observed cells in the n -th generation distinguishing their type: $Z_n^0 = |\mathbb{G}_n^* \cap 2\mathbb{N}|$, $Z_n^1 = |\mathbb{G}_n^* \cap (2\mathbb{N} + 1)|$, and we set, for all $n \geq 1$, $\mathbf{Z}_n = (Z_n^0, Z_n^1)$. Note that for $i \in \{0, 1\}$ and $n \geq 1$ one has $Z_n^i = \sum_{k \in \mathbb{G}_{n-1}} \delta_{2k+i}$.

3. Asymmetry in the lineage

We now describe the mechanism generating the observation process (δ_k) and recall some classical assumptions taken from Harris (1963). We define the genealogy of the cells by a two-type GW process (\mathbf{Z}_n) . All cells reproduce independently and with a reproduction law depending only on their type. For a mother cell of type $i \in \{0, 1\}$, we denote by $p^{(i)}(j_0, j_1)$ the probability that it has j_0 daughter of type 0 and j_1 daughter of type 1. For the cell division process we are interested in, one clearly has $p^{(i)}(0, 0) + p^{(i)}(1, 0) + p^{(i)}(0, 1) + p^{(i)}(1, 1) = 1$. The reproduction laws have moments of all orders, and we can thus define the descendant matrix as the 2×2 matrix $\mathbf{P} = (p_{ij})_{0 \leq i, j \leq 1}$, where $p_{i0} = p^{(i)}(1, 0) + p^{(i)}(1, 1)$ and $p_{i1} = p^{(i)}(0, 1) + p^{(i)}(1, 1)$, for $i \in \{0, 1\}$ i.e. p_{ij} is the expected number of descendants of type j of a cell of type i . We make the following main assumption.

(AO) All entries of the matrix \mathbf{P} are positive and its dominant eigenvalue π satisfies $\pi > 1$.

In this case, there exist component-wise positive left eigenvectors for π . Let $\mathbf{z} = (z^0, z^1)$ be the one satisfying $z^0 + z^1 = 1$. Assumption (AO) means that (\mathbf{Z}_n) is super-critical and ensures that extinction is not almost sure: $\mathbb{P}(\mathcal{E}) < 1$. On $\bar{\mathcal{E}}$, $|\mathbb{T}_n^*|^{-1} \sum_{i=1}^n Z_i^i$ converges to z^i , meaning that z^i is the asymptotic proportion of cells of type i . Our context is very specific because the information given by (δ_k) is more precise than the one given by (\mathbf{Z}_n) used in the literature, see Guttorp (1991). Empiric estimators of the reproduction probabilities using data up to the n -th generation are, for i, j_0, j_1 in $\{0, 1\}$

$$\hat{p}_n^{(i)}(j_0, j_1) = \frac{\sum_{k \in \mathbb{T}_{n-2}} \delta_{2k+i} \phi_{j_0}(\delta_{2(2k+i)}) \phi_{j_1}(\delta_{2(2k+i)+1})}{\sum_{k \in \mathbb{T}_{n-2}} \delta_{2k+i}},$$

where $\phi_0(x) = 1 - x$, $\phi_1(x) = x$, if the denominators are nonzero (zero otherwise). The strong consistency is readily obtained on the non-extinction set $\bar{\mathcal{E}}$ by martingale methods.

Lemma 3.1. Under (AO) and for all i, j_0 and j_1 in $\{0, 1\}$, one has

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \hat{p}_n^{(i)}(j_0, j_1) = p^{(i)}(j_0, j_1) \mathbb{1}_{\bar{\mathcal{E}}} \text{ a.s.}$$

Set $\mathbf{p}^{(i)} = (p^{(i)}(0, 0), p^{(i)}(1, 0), p^{(i)}(0, 1), p^{(i)}(1, 1))^t$ the vector of the 4 reproduction probabilities for a mother of type i , $\mathbf{p} = ((\mathbf{p}^{(0)})^t, (\mathbf{p}^{(1)})^t)^t$ the vector of all 8 reproduction probabilities and $\hat{\mathbf{p}}_n = (\hat{p}_n^{(0)}(0, 0), \dots, \hat{p}_n^{(1)}(1, 1))^t$ its estimator. As $\mathbb{P}(\bar{\mathcal{E}}) \neq 0$, we define the conditional probability $\mathbb{P}_{\bar{\mathcal{E}}}$ by $\mathbb{P}_{\bar{\mathcal{E}}}(A) = \mathbb{P}(A \cap \bar{\mathcal{E}}) / \mathbb{P}(\bar{\mathcal{E}})$ for all events A .

Theorem 3.2. Under assumption (AO), we have the convergence $\sqrt{|\mathbb{T}_{n-1}^*|}(\hat{\mathbf{p}}_n - \mathbf{p}) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \mathbf{V})$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$, with $\mathbf{V} = \begin{pmatrix} \mathbf{v}^0/z^0 & 0 \\ 0 & \mathbf{v}^1/z^1 \end{pmatrix}$ and for all i in $\{0, 1\}$, $\mathbf{V}^i = \mathbf{W}^i - \mathbf{p}^{(i)}(\mathbf{p}^{(i)})^t$, \mathbf{W}^i is a 4×4 matrix with the entries of $\mathbf{p}^{(i)}$ on the diagonal and 0 elsewhere.

Proof. For all $n \geq 2$, and $q \geq 1$, set $\mathbf{M}_q^{(n)} = ((\mathbf{M}_q^{0(n)})^t, (\mathbf{M}_q^{1(n)})^t)^t$, with

$$\mathbf{M}_q^{(n)} = \frac{1}{\sqrt{|\mathbb{T}_{n-1}^*|}} \sum_{k=1}^q \begin{pmatrix} \delta_{2k+i}((1 - \delta_{2(2k+i)})(1 - \delta_{2(2k+i)+1}) - p^{(i)}(0, 0)) \\ \delta_{2k+i}(\delta_{2(2k+i)}(1 - \delta_{2(2k+i)+1}) - p^{(i)}(1, 0)) \\ \delta_{2k+i}((1 - \delta_{2(2k+i)})\delta_{2(2k+i)+1} - p^{(i)}(0, 1)) \\ \delta_{2k+i}(\delta_{2(2k+i)}\delta_{2(2k+i)+1} - p^{(i)}(1, 1)) \end{pmatrix}.$$

Let (\mathcal{G}_q) be the filtration of cousin cells: $\mathcal{G}_0 = \sigma\{\delta_1, \delta_2, \delta_3\}$ and for all $q \geq 1$, $\mathcal{G}_q = \mathcal{G}_{q-1} \vee \sigma\{\delta_{4q}, \delta_{4q+1}, \delta_{4q+2}, \delta_{4q+3}\}$. For all $n \geq 2$, $(\mathbf{M}_q^{(n)})$ is a (\mathcal{G}_q) -martingale. We apply Theorem 3.II.10 of Duflo (1997) with the stopping times $\nu_n = |\mathbb{T}_{n-2}|$. The $\mathbb{P}_{\bar{\mathcal{E}}}$ a.s. limit of the increasing process is

$$\langle \mathbf{M}^{(n)} \rangle_{\nu_n} = \frac{1}{|\mathbb{T}_{n-1}^*|} \begin{pmatrix} \sum_{k \in \mathbb{T}_{n-2}} \delta_{2k} \mathbf{V}^0 & 0 \\ 0 & \sum_{k \in \mathbb{T}_{n-2}} \delta_{2k+1} \mathbf{V}^1 \end{pmatrix} \xrightarrow{n \rightarrow \infty} \mathbf{G} = \begin{pmatrix} z^0 \mathbf{V}^0 & 0 \\ 0 & z^1 \mathbf{V}^1 \end{pmatrix}.$$

In addition, the Lindeberg condition holds as the δ_k have finite moments of all orders. Thus, we obtain the convergence $M_{|\mathbb{T}_{n-2}|}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{G})$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. On the other hand, $\Delta_{n-1}^{-1} |\mathbb{T}_{n-1}^*| M_{|\mathbb{T}_{n-2}|}^{(n)} = \sqrt{|\mathbb{T}_{n-1}^*|} (\hat{\mathbf{p}}_n - \mathbf{p})$, with

$$\Delta_n = \begin{pmatrix} \sum_{\ell=1}^n Z_\ell^0 \mathbf{I}_4 & 0 \\ 0 & \sum_{\ell=1}^n Z_\ell^1 \mathbf{I}_4 \end{pmatrix},$$

and \mathbf{I}_4 is the identity matrix of size 4. As $|\mathbb{T}_n^*|^{-1} \sum_{\ell=1}^n Z_\ell^i$ converges a.s. to z^i on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$, we have the asymptotic normality, using Slutsky's lemma. \square

We now derive Wald's test for the asymmetry of the means of the reproduction laws. Set $m = (p^{(0)}(1, 0) + p^{(0)}(0, 1) + 2p^{(0)}(1, 1)) - (p^{(1)}(1, 0) + p^{(1)}(0, 1) + 2p^{(1)}(1, 1))$ the difference of the means of the types 0 and 1 reproduction laws and \hat{m}_n its empirical estimator. Set $\mathbf{H}_0^{\text{GW}}: m = 0$ the symmetry hypothesis and $\mathbf{H}_1^{\text{GW}}: m \neq 0$. Let $(Y_n^{\text{GW}})^2$ be the test statistic defined by $Y_n^{\text{GW}} = |\mathbb{T}_{n-1}^*|^{1/2} (\hat{\Delta}_n^{\text{GW}})^{-1/2} \hat{m}_n$, where $\hat{\Delta}_n^{\text{GW}} = \mathbf{d} \mathbf{g} \mathbf{w}^t \hat{\mathbf{V}}_n \mathbf{d} \mathbf{g} \mathbf{w}$, $\mathbf{d} \mathbf{g} \mathbf{w} = (0, 1, 1, 2, 0 - 1, -1, -2)^t$, and $\hat{\mathbf{V}}_n$ is the empirical version of \mathbf{V} , where z^i is replaced by $|\mathbb{T}_n^*|^{-1} \sum_{i=1}^n Z_i^i$ and the $p^{(i)}(j_0, j_1)$ are replaced by $\hat{p}_n^{(i)}(j_0, j_1)$. Thanks to Lemma 3.1, $\hat{\mathbf{V}}_n$ converges a.s. to \mathbf{V} and the test statistic has the following asymptotic properties.

Theorem 3.3. Under assumption (AO) and the null hypothesis \mathbf{H}_0^{GW} , one has $(Y_n^{\text{GW}})^2 \xrightarrow{\mathcal{L}} \chi^2(1)$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^{GW} , one has $\lim_{n \rightarrow \infty} (Y_n^{\text{GW}})^2 = +\infty$ a.s. on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$.

Proof. Let g be the function defined from \mathbb{R}^8 onto \mathbb{R} by $g(x_1, \dots, x_8) = (x_2 + x_3 + 2x_4) - (x_6 + x_7 + 2x_8)$, so that $\hat{m}_n - m = g(\hat{\mathbf{p}}_n) - g(\mathbf{p})$, and $\mathbf{d} \mathbf{g} \mathbf{w}$ is the gradient of g . Theorem 3.2 yields $\sqrt{|\mathbb{T}_{n-1}^*|} (g(\hat{\mathbf{p}}_n) - g(\mathbf{p})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Delta^{\text{GW}})$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$, with $\Delta^{\text{GW}} = \mathbf{d} \mathbf{g} \mathbf{w}^t \mathbf{V} \mathbf{d} \mathbf{g} \mathbf{w}$. Under \mathbf{H}_0^{GW} , $g(\mathbf{p}) = m = 0$, so that $|\mathbb{T}_{n-1}^*| (\Delta^{\text{GW}})^{-1} g(\hat{\mathbf{p}}_n)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. One then uses Slutsky's lemma to replace Δ^{GW} by $\hat{\Delta}_n^{\text{GW}}$. Under \mathbf{H}_1^{GW} , since $Y_n^{\text{GW}} = |\mathbb{T}_{n-1}^*|^{1/2} (\hat{\Delta}_n^{\text{GW}})^{-1/2} \hat{m}_n$ and \hat{m}_n converges to $m \neq 0$, Y_n^{GW} tends to infinity a.s. on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. \square

4. Asymmetry in cell's characteristic

We turn to the asymmetry of the BAR model with missing data. Assume that $\mathbb{E}[X_k^8] < \infty$ and $0 < \max(|b|, |d|) < 1$. Denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration associated with the BAR process: $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$.

(AN.1) One has $\sup_{n \geq 0} \sup_{k \in \mathbb{G}_{n+1}} \mathbb{E}[\varepsilon_k^8 | \mathcal{F}_n] < \infty$ a.s., $\forall n \geq 0$ and $k \in \mathbb{G}_{n+1}$, $\mathbb{E}[\varepsilon_k | \mathcal{F}_n] = 0$, $\mathbb{E}[\varepsilon_k^2 | \mathcal{F}_n] = \sigma^2$ a.s., $\forall k \in \mathbb{G}_n$, $\mathbb{E}[\varepsilon_{2k} \varepsilon_{2k+1} | \mathcal{F}_n] = \rho$ a.s.

(AN.2) $\forall n \geq 0$, $\{(\varepsilon_{2k}, \varepsilon_{2k+1})\}_{k \in \mathbb{G}_n}$ are conditionally independent given \mathcal{F}_n .

(AI) The sequence (δ_k) is independent from the sequences (X_k) and (ε_k) .

The least-squares estimator of $\theta = (a, b, c, d)^t$ is given for all $n \geq 1$ by $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n)^t$ with

$$\hat{\theta}_n = \Sigma_n^{-1} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{2k} X_{2k} \\ \delta_{2k} X_k X_{2k} \\ \delta_{2k+1} X_{2k+1} \\ \delta_{2k+1} X_k X_{2k+1} \end{pmatrix}, \quad \Sigma_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix},$$

$$\mathbf{S}_n^i = \sum_{k \in \mathbb{T}_n} \delta_{2k+i} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}, \quad \mathbf{S}_n^{0,1} = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

Denote also $\mathbf{L}^0, \mathbf{L}^1, \mathbf{L}^{0,1}$ their a.s. limits: $\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathbf{S}_n^i / |\mathbb{T}_n^*| = \mathbb{1}_{\bar{\mathcal{E}}} \mathbf{L}^i$, $\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathbf{S}_n^{0,1} / |\mathbb{T}_n^*| = \mathbb{1}_{\bar{\mathcal{E}}} \mathbf{L}^{0,1}$, see Proposition 4.2 of de Saporta et al. (2011). We now recall Theorems 3.2 and 3.4 of de Saporta et al. (2011).

Table 1

Proportions of p -values under the 0.05, 0.01 and 0.001 thresholds of the asymmetry tests for the means of the GW process (1000 replicas) $\mathbf{p}^{(0)} = (0.04, 0.08, 0.08, 0.8)$ (under \mathbf{H}_1), $\mathbf{p}^{(1)} = (0.15, 0.08, 0.08, 0.69)$.

Generation	Under \mathbf{H}_0^{GW}			Under \mathbf{H}_1^{GW}		
	$p < 0.05$	$p < 0.01$	$p < 0.001$	$p < 0.05$	$p < 0.01$	$p < 0.001$
7	6.4	1.9	0.3	27.8	11.8	03.6
8	5.6	1.4	0.3	44.2	22.2	07.6
9	5.5	1.1	0.3	58.6	38.5	17.0
10	5.7	1.5	0.2	79.4	60.8	35.9
11	4.8	1.0	0.1	93.1	82.0	64.2

Theorem 4.1. Under (AN.1–2), (AO) and (AI), the estimator $\hat{\theta}_n$ is strongly consistent $\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \hat{\theta}_n = \theta \mathbb{1}_{\bar{\mathcal{E}}}$ a.s. In addition, we have the asymptotic normality $\sqrt{|\mathbb{T}_{n-1}^*|}(\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Sigma^{-1} \Gamma \Sigma^{-1})$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$, where

$$\Sigma = \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix}, \quad \text{and} \quad \Gamma = \begin{pmatrix} \sigma^2 \mathbf{L}^0 & \rho \mathbf{L}^{0,1} \\ \rho \mathbf{L}^{0,1} & \sigma^2 \mathbf{L}^1 \end{pmatrix}.$$

We now propose two different asymmetry tests. The first one compares the couples (a, b) and (c, d) . Set $\mathbf{H}_0^c: (a, b) = (c, d)$ the symmetry hypothesis and $\mathbf{H}_1^c: (a, b) \neq (c, d)$. Let $(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c$ be the test statistic defined by $\mathbf{Y}_n^c = |\mathbb{T}_{n-1}^*|^{1/2} (\hat{\mathbf{A}}_n^c)^{-1/2} (\hat{a}_n - \hat{c}_n, \hat{b}_n - \hat{d}_n)^t$, with $\hat{\mathbf{A}}_n^c = |\mathbb{T}_{n-1}^*|^2 \mathbf{dgc}^t \Sigma_{n-1}^{-1} \hat{\Gamma}_{n-1} \Sigma_{n-1}^{-1} \mathbf{dgc}$,

$$\mathbf{dgc} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}^t, \quad \hat{\Gamma}_n = \frac{1}{|\mathbb{T}_n^*|} \begin{pmatrix} \hat{\sigma}_{n+1}^2 \mathbf{S}_n^{0,0} & \hat{\rho}_{n+1} \mathbf{S}_n^{0,1} \\ \hat{\rho}_{n+1} \mathbf{S}_n^{0,1} & \hat{\sigma}_{n+1}^2 \mathbf{S}_n^{1,1} \end{pmatrix},$$

$\hat{\sigma}_n^2 = |\mathbb{T}_n^*|^{-1} \sum_{k \in \mathbb{T}_{n-1}^*} (\hat{\varepsilon}_{2k}^2 + \hat{\varepsilon}_{2k+1}^2)$, $\hat{\rho}_n = |\mathbb{T}_{n-1}^{*01}|^{-1} \sum_{k \in \mathbb{T}_{n-1}^*} \hat{\varepsilon}_{2k} \hat{\varepsilon}_{2k+1}$, $\mathbb{T}_n^{*01} = \{k \in \mathbb{T}_n : \delta_{2k} \delta_{2k+1} = 1\}$ and for all $k \in \mathbb{G}_n$, $\hat{\varepsilon}_{2k} = \delta_{2k} (X_{2k} - \hat{a}_n - \hat{b}_n X_k)$, $\hat{\varepsilon}_{2k+1} = \delta_{2k+1} (X_{2k+1} - \hat{c}_n - \hat{d}_n X_k)$.

Theorem 4.2. Under assumptions (AN.1–2), (AO), (AI) and the null hypothesis \mathbf{H}_0^c , one has $(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c \xrightarrow{\mathcal{L}} \chi^2(2)$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^c , one has $\lim_{n \rightarrow \infty} \|\mathbf{Y}_n^c\|^2 = +\infty$ a.s. on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$.

Proof. We mimic the proof of Theorem 3.3 with g the function defined from \mathbb{R}^4 onto \mathbb{R}^2 by $g(x_1, x_2, x_3, x_4) = (x_1 - x_3, x_2 - x_4)^t$. \square

Our second test compares the fixed points $a/(1 - b)$ and $c/(1 - d)$, which are the asymptotic means of X_{2k} and X_{2k+1} respectively. Set $\mathbf{H}_0^f: a/(1 - b) = c/(1 - d)$ the symmetry hypothesis and $\mathbf{H}_1^f: a/(1 - b) \neq c/(1 - d)$. Let $(Y_n^f)^2$ be the test statistic defined by $Y_n^f = |\mathbb{T}_{n-1}^*|^{1/2} (\hat{\mathbf{A}}_n^f)^{-1/2} (\hat{a}_n/(1 - \hat{b}_n) - \hat{c}_n/(1 - \hat{d}_n))$, where $\hat{\mathbf{A}}_n^f = |\mathbb{T}_{n-1}^*|^2 \mathbf{dgf}^t \Sigma_{n-1}^{-1} \hat{\Gamma}_{n-1} \Sigma_{n-1}^{-1} \mathbf{dgf}$, $\mathbf{dgf} = (1/(1 - b), a/(1 - b)^2, -1/(1 - d), -c/(1 - d)^2)^t$.

Theorem 4.3. Under assumptions (AN.1–2), (AO), (AI) and the null hypothesis \mathbf{H}_0^f , one has $(Y_n^f)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$ on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^f , one has $\lim_{n \rightarrow \infty} (Y_n^f)^2 = +\infty$ a.s. on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$.

Proof. We mimic again the proof of Theorem 3.3 with g the function defined from \mathbb{R}^4 onto \mathbb{R} by $g(x_1, x_2, x_3, x_4) = (x_1/(1 - x_2) - x_3/(1 - x_4))^t$. \square

5. Application to simulated data

We now study the behavior of our tests on simulated data. For each test, we compute, in function of the generation n and for different thresholds, the proportion of rejections under hypotheses \mathbf{H}_0 and \mathbf{H}_1 , the latter being an indicator of the power of the test. Proportions are computed on a sample of 1000 replicated trees. In Table 1, the observed proportions of p -values under the given thresholds are close to the expected proportions of rejection under \mathbf{H}_0^{GW} suggesting that the asymptotic law of the statistic $(\mathbf{Y}_n^{GW})^2$ is available by generation 8. Under \mathbf{H}_1^{GW} , the power of the test increases from 27.8% for generation 7 to 93.1% for generation 11 for a risk of type 1 fixed at 0.05. In Table 2, the observed proportions of p -values under the given thresholds, are close to the expected proportions of rejection under \mathbf{H}_0^c suggesting that the asymptotic law of the statistic $\|\mathbf{Y}_n^c\|^2$ is also available at generation 8. Under \mathbf{H}_1^c , the power of the test increases from 37.4% for generation 7 to 95.7% for generation 11 for a risk of type 1 fixed at 0.05. In Table 3, the observed proportions go away from the expected ones under \mathbf{H}_0^f , suggesting that the asymptotic law of the statistic is not reached before the 10th generation. The power is also weak until the 10th generation.

Table 2

Proportions of p -values under the 0.05, 0.01 and 0.001 thresholds of the asymmetry test for the parameters of the BAR process (1000 replicas) $a = b = 0.5$ (under H_1), $c = 0.5$; $d = 0.4$.

Gen	Under H_0^c			Under H_1^c		
	$p < 0.05$	$p < 0.01$	$p < 0.001$	$p < 0.05$	$p < 0.01$	$p < 0.001$
7	6.6	2.2	0.6	37.4	19.7	08.0
8	5.5	1.5	0.3	53.6	31.0	14.6
9	5.5	1.3	0.3	71.1	52.3	30.3
10	6.3	1.2	0.1	86.8	75.5	56.1
11	5.9	0.6	0.1	95.7	90.8	81.4

Table 3

Proportions of p -values under the 0.05, 0.01 and 0.001 thresholds of the asymmetry test for the fixed points of the BAR process (1000 replicas) $a = b = 0.5$ (under H_1), $c = 0.5$; $d = 0.4$.

Gen	Under H_0^f			Under H_1^f		
	$p < 0.05$	$p < 0.01$	$p < 0.001$	$p < 0.05$	$p < 0.01$	$p < 0.001$
7	2.2	0.7	0	23.1	07.4	01.4
8	3.3	0.5	0.1	41.3	20.5	06.1
9	3.8	0.5	0	64.6	41.6	18.6
10	4.7	0.8	0	82.9	68.1	46.3
11	5.5	0.7	0.1	94.5	88.5	74.5

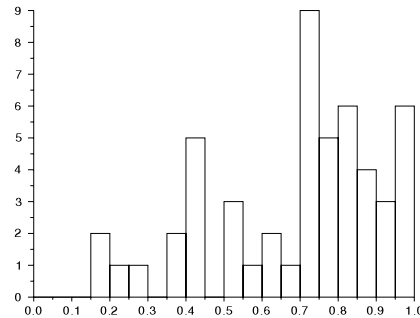


Fig. 1. Histogram of the 51 p -values of the test H_0^{GW} .

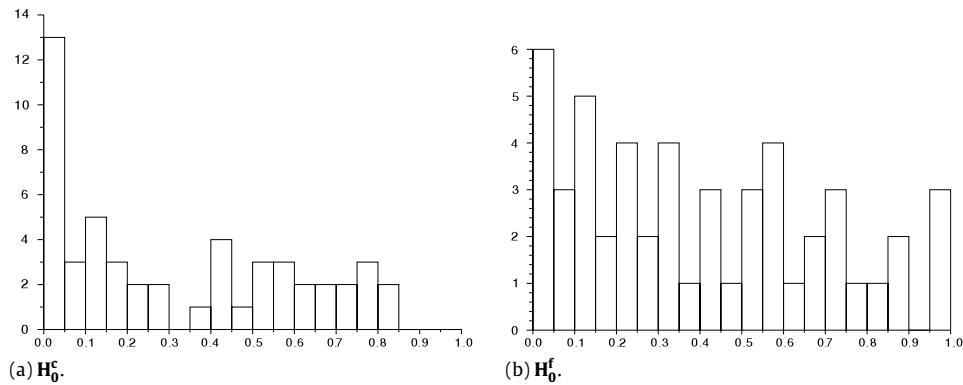


Fig. 2. Histogram of the 51 p -values of the tests H_0^c and H_0^f .

6. Application to real data: aging detection of Escherichia coli

To study aging in the single cell organism *E. coli*, Stewart et al. (2005) filmed 94 colonies of dividing cells, determining the complete lineage and the growth rate of each cell. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell inherits an old end or *pole* from its mother, and creates a new pole. Therefore, each cell has a *type*: old pole or new pole inducing asymmetry in the cell division. Stewart et al. (2005) propose a statistical study of the mean genealogy and pair-wise comparison of sister cells assuming their independence which is not verified in the lineage. We ran our tests of the null hypotheses H_0^{GW} , H_0^c and H_0^f on the 51 data sets issued of the 94 colonies containing at least eight generations. Fig. 1 shows that the hypothesis of equality of the expected number of observed offspring between two sisters is not rejected whatever the data set. This result is not surprising: data are missing most frequently because the cells were out of the

range of the camera. The null hypotheses of the two tests are rejected for one set in four for H_0^c and for one in eight for H_0^f (see Fig. 2). A global conclusion on the asymmetry between the old pole and new pole cells is not easy. Regarding the simulation results in Tables 2 and 3, this lack of evidence is probably due to a low power of the tests at generations 8 and 9.

References

- Bercu, B., de Saporta, B., Gégout-Petit, A., 2009. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* 14 (87), 2492–2526.
- Cowan, R., Staudte, R.G., 1986. The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42, 769–783.
- Delmas, J.F., Marsalle, L., 2010. Detection of cellular ageing in a Galton–Watson process. *Stochastic Process. Appl.* 120, 2495–2519.
- de Saporta, B., Gégout-Petit, A., Marsalle, L., 2011. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Stat.* 5, 1313–1353.
- Duflo, M., 1997. Random Iterative Models. In: *Applications of Mathematics*, vol. 34. Springer-Verlag, Berlin.
- Guttorp, P., 1991. Statistical Inference for Branching Processes. In: *Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics*, John Wiley & Sons Inc., New York.
- Guyon, J., 2007. Limit theorems for bifurcating Markov chains. Application to the detection of cellular ageing. *Ann. Appl. Probab.* 17, 1538–1569.
- Harris, T.E., 1963. The Theory of Branching Processes. In: *Die Grundlehren der Mathematischen Wissenschaften*, Bd. 119. Springer-Verlag, Berlin.
- Hwang, S.Y., Basawa, I.V., 2009. Branching Markov processes and related asymptotics. *J. Multivariate Anal.* 100, 1155–1167.
- Hwang, S.Y., Basawa, I.V., 2011. Asymptotic optimal inference for multivariate branching-Markov processes via martingale estimating functions and mixed normality. *J. Multivariate Anal.* 102, 1018–1031.
- Stewart, E., Madden, R., Paul, G., Taddei, F., 2005. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* 3, e45.

- 1.4 [dSGP⁺13a] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Random coefficients bifurcating autoregressive processes. En révision pour ESAIM PS, arXiv :1205.3658, 2013.

Random coefficient bifurcating autoregressive processes

Benoîte de Saporta, Anne Gégout Petit, Laurence Marsalle

Abstract

This paper presents a model of asymmetric bifurcating autoregressive process with random coefficients. We couple this model with a Galton Watson tree to take into account possibly missing observations. We propose least-squares estimators for the various parameters of the model and prove their consistency with a convergence rate, and their asymptotic normality. We use both the bifurcating Markov chain and martingale approaches and derive new important general results in both these frameworks.

1 Introduction

In the 80's, Cowan and Staudte [1] introduced Bifurcating Autoregressive processes (BAR) as a parametric model to study cell lineage data. A quantitative characteristic of the cells (e.g. growth rate, age at division) is recorded over several generations descended from an initial cell, keeping track of the genealogy to study inherited effects. As a cell usually gives birth to two offspring by division, such genealogies are naturally structured as binary trees. BAR processes are thus a generalization of autoregressive processes (AR) to this binary tree structure, by modeling each line of descent as a first order AR process, allowing the environmental effects on sister cells to be correlated. Statistical inference for the parameters of BAR processes has been widely studied, either based on the observation of a single tree growing to infinity [1, 2, 3, 4] or on a large number of small independent trees [5, 6]. See also [7, 8] for processes indexed by general trees.

Various extensions of the original model have been proposed, e.g. non gaussian noise sequence [9, 10], higher order AR [3, 10] or moving average AR [6]. Since 2005, evidence of asymmetry in cell division has been established by biologists [11] and an asymmetric BAR model has been introduced by Guyon [12] where the coefficients of the AR processes of sister cells are allowed to be different. This model was further extended to higher order AR [13], to take missing data into account [14, 15, 16] and with parasite infection [17].

To our best knowledge, only two papers [18] and [19] deal with random coefficient BAR processes. In the former by Bui and Huggins it is explained that random coefficients BAR processes can account for observations that do not fit the usual BAR model. For instance, the extra randomness can model irregularities in nutrient concentrations in the media in which the cells are grown. Other evidence for the need

of richer models can be found e.g. in [20]. In this paper, we propose a new model for random coefficient BAR processes (R-BAR). It is more general than that of Bui and Huggins, as the random variables are not supposed to be Gaussian, they may not have moments of all order and correlation between all the sources of randomness are allowed. Moreover, we propose an asymmetric model in the continuance of [12, 13, 14, 15, 16, 19] in the context of missing data. Indeed, experimental data are often incomplete and it is important to take this phenomenon into account for the inference. As in [14, 15] we model the structure of available data by a Galton Watson tree, instead of a complete binary tree. Our model is close to that developed in [19], but the assumptions on the noise process are different as we allow correlation between the two sources of randomness but require higher moments because of the missing data and because we do not use a weighted estimator. The main difference is that the model in [19] is fully observed, whereas ours allows for missing observations.

Our approach for the inference is also different from [18, 19]. As we cannot use maximum likelihood estimation, we propose modified least squares estimators as in [21]. In [18], inference is based on an asymptotically infinite number of small replicated trees. Here, as in [19], we consider one single tree growing to infinity but our least squares estimator is not weighted. The originality of our approach is that it combines the bifurcating Markov chain and martingale approaches. Bifurcating Markov chains (BMC) were introduced in [12] on complete binary trees and further developed in [14] in the context of missing data on Galton Watson trees. BAR models can be seen as a special case of BMC. This interpretation allows us to establish the convergence of our estimators. A by-product of our procedure is a new general result for BMC on Galton Watson trees. Indeed, in [12, 14] the driven noise sequence is assumed to have moments of all order. Here, we establish new laws of large numbers for polynomial functions of the BMC where the noise sequence only has moments up to a given order. The strong law of large numbers [22] and the central limit theorem [23, 24, 22] for martingales have been previously used in the context of BAR processes [9, 10, 4] and adapted to special cases of martingales on binary trees [13, 15, 16, 19]. In this paper, we establish a general law of large numbers for square integrable martingales on Galton Watson binary trees. This result is applied to our R-BAR model to obtain sharp convergence rates and a quadratic strong law for our estimators.

The paper is organized as follows. In Section 2, we give the precise definition of our R-BAR model on a Galton Watson tree and state our main assumptions. In Section 3, we give modified least squares estimators and state the convergence results we obtained: consistency with convergence rate and asymptotic normality. In Section 4, we recall the BMC framework, prove a new law of large numbers under limited moment conditions and apply it to our R-BAR model to derive the consistency of our estimators. In Section 5 we establish a new general law of large numbers for square integrable martingales on Galton Watson trees and use it to derive convergence rates and quadratic strong laws for our estimators. In Section 6 we establish the asymptotic normality by using central limit theorems for martingales. Finally in Section 7 we apply our estimation procedure to *E. coli* data of [11].

2 Model

In the sequel, all random variables are defined on the probability state space $(\Omega, \mathcal{A}, \mathbb{P})$. As in the previous literature, we use the index 1 for the original cell, and the two offspring of cell k are labelled $2k$ and $2k + 1$. Consider the first-order asymmetric random coefficients bifurcating autoregressive process (R-BAR) given, for all $k \geq 1$, by

$$\begin{cases} X_{2k} &= (b_{2k} + \eta_{2k})X_k + (a_{2k} + \varepsilon_{2k}), \\ X_{2k+1} &= (b_{2k+1} + \eta_{2k+1})X_k + (a_{2k+1} + \varepsilon_{2k+1}), \end{cases} \quad (2.1)$$

where for all $k \geq 1$, one has $a_{2k} = a$, $b_{2k} = b$, $a_{2k+1} = c$ and $b_{2k+1} = d$. The initial state X_1 is the characteristic of the original ancestor while the sequence $(\varepsilon_{2k}, \eta_{2k}, \varepsilon_{2k+1}, \eta_{2k+1})_{n \geq 1}$ is the driven noise of the process, and the parameter (a, b, c, d) belongs to \mathbb{R}^4 . One can see this R-BAR process as a random-coefficient first-order autoregressive process on a binary tree, where each vertex represents an individual or cell, vertex 1 being the original ancestor. For all $n \geq 1$, denote the n -th generation by \mathbb{G}_n and the sub-tree of all individuals from the original individual up to the n -th generation by \mathbb{T}_n :

$$\mathbb{G}_n = \{2^n, 2^n + 1, \dots, 2^{n+1} - 1\}, \quad \mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$$

In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the original ancestor. Finally, denote by \mathbb{T} the complete tree. Note that the cardinality $|\mathbb{G}_n|$ of \mathbb{G}_n is 2^n while that of \mathbb{T}_n is $|\mathbb{T}_n| = 2^{n+1} - 1$. In the sequel, we shall make use of the following hypotheses.

(H.1) The sequence $(\varepsilon_{2k}, \eta_{2k}, \varepsilon_{2k+1}, \eta_{2k+1})_{k \geq 1}$ is independent and identically distributed. It is also independent from X_1 .

(H.2) The random variables $\varepsilon_2, \eta_2, \varepsilon_3, \eta_3$ and X_1 have moments of all order up to 4γ , for some $\gamma \geq 1$. One has

$$\begin{aligned} \mathbb{E}[\varepsilon_2] &= \mathbb{E}[\varepsilon_3] = 0, & \mathbb{E}[\varepsilon_2^2] &= \mathbb{E}[\varepsilon_3^2] = \sigma_\varepsilon^2 > 0, & \mathbb{E}[\varepsilon_2 \varepsilon_3] &= \rho_\varepsilon, \\ \mathbb{E}[\eta_2] &= \mathbb{E}[\eta_3] = 0, & \mathbb{E}[\eta_2^2] &= \mathbb{E}[\eta_3^2] = \sigma_\eta^2 > 0, & \mathbb{E}[\eta_2 \eta_3] &= \rho_\eta, \\ \mathbb{E}[\varepsilon_{2+i} \eta_{2+j}] &= \rho_{ij}, & \text{for } (i, j) &\in \{0, 1\}, & \text{and } \rho &= \frac{1}{2}(\rho_{01} + \rho_{10}). \end{aligned}$$

In addition, for all p, q, r, s such that $p + q + r + s \leq 4\gamma$ denote

$$\mathbb{E}[\varepsilon_2^p \eta_2^q \varepsilon_3^r \eta_3^s] = \vartheta(p, q, r, s).$$

When dealing with the biological issue of cell division, it may happen that a lineage is incomplete. Indeed, cells may die or measurements may be impossible or faulty on some cells. Taking into account such a phenomenon, we introduce the observation process, $(\delta_k)_{k \in \mathbb{T}}$. Basically, $\delta_k = 1$ if cell k is observed, $\delta_k = 0$ otherwise. We use the same framework as in [14], and not the more general introduced in [15]. We set $\delta_1 = 1$ and define the whole sequence through the following equalities:

$$\delta_{2k} = \delta_k \xi_k^0 \quad \text{and} \quad \delta_{2k+1} = \delta_k \xi_k^1, \quad (2.2)$$

where the sequence $(\boldsymbol{\xi}_k = (\xi_k^0, \xi_k^1))_{k \in \mathbb{T}}$ is a sequence of independent identically distributed random vectors of $\{0, 1\}^2$ with generating function

$$\mathbb{E}[s_0^{\xi_0^0} s_1^{\xi_0^1}] = (1 - p_0 - p_1 - p_{01}) + p_0 s_0 + p_1 s_1 + p_{01} s_0 s_1.$$

We also suppose independence between the observation and state processes.

(H.3) The sequence $(\boldsymbol{\xi}_k)_{k \in \mathbb{T}}$ is independent from $(\varepsilon_{2k}, \eta_{2k}, \varepsilon_{2k+1}, \eta_{2k+1})_{k \in \mathbb{T}}$ and from X_1 .

Notice that the process $(\delta_k)_{k \in \mathbb{T}}$ takes its values in $\{0, 1\}$, and that if $k \in \mathbb{T}$ is such that $\delta_k = 0$, then $\delta_{2^n k + i} = 0$, for all $i \in \{0, \dots, 2^n - 1\}$ and all $n \geq 1$. So to speak, if individual k is not observed, all its descendants are also missing. We now define the sets of observed data

$$\mathbb{G}_n^* = \{k \in \mathbb{G}_n : \delta_k = 1\} \quad \text{and} \quad \mathbb{T}_n^* = \{k \in \mathbb{T}_n : \delta_k = 1\} = \cup_{\ell=0}^n \mathbb{G}_\ell^*.$$

Thanks to the i.i.d. property of $(\boldsymbol{\xi}_k)$, the sequence of cardinalities $(|\mathbb{G}_n^*|)_{n \geq 0}$ is a Galton-Watson (GW) process with reproduction generating function

$$z \longmapsto (1 - p_0 - p_1 - p_{01}) + (p_0 + p_1)z + p_{01}z^2,$$

and mean $m = 2p_{01} + p_0 + p_1$. According to the position of m of with respect to 1, it is well known that the population extincts a.s. or not. More precisely, if $m \leq 1$ then we have extinction almost surely, in the sense that $\mathbb{P}(\cup_{n \geq 0} \{|\mathbb{G}_n^*| = 0\}) = 1$. But if $m > 1$, there is a positive probability of survival of the population: $\mathbb{P}(\cap_{n \geq 0} \{|\mathbb{G}_n^*| > 0\}) > 0$. This latter case is called the super-critical case, and we assume that we are in that case.

(H.4) The mean of the reproduction law is greater than 1: $m > 1$.

On the non-extinction set, the growth of the population is exponential. There exists some non-negative square integrable random variable W such that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{G}_n^*|}{m^n} = W \text{ a.s.}, \quad \text{and} \quad \{W > 0\} = \cap_{n \geq 0} \{|\mathbb{G}_n^*| > 0\} \text{ a.s.} \quad (2.3)$$

This immediately entails that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n^*|}{m^n} = W \times \frac{m}{m-1} \text{ a.s.} \quad (2.4)$$

We will denote \mathcal{E} the extinction set $\mathcal{E} = \cup_{n \geq 0} \{|\mathbb{G}_n^*| = 0\}$ and $\overline{\mathcal{E}}$ its complementary set. Note that under assumption **(H.4)**, one has $\mathbb{P}(\overline{\mathcal{E}}) > 0$. We need one more assumption combining the R-BAR and GW processes.

(H.5) There exist $1 \leq \kappa \leq \gamma$ such that

$$\frac{p_0 + p_{01}}{m} \mathbb{E}[(b + \eta_2)^{4\kappa}] + \frac{p_1 + p_{01}}{m} \mathbb{E}[(d + \eta_3)^{4\kappa}] < 1.$$

This is the analogous of the usual assumption $\max\{|b|, |d|\} < 1$ in the case of fixed coefficients. The assumption above is slightly weaker: in the fully observed case and when $\eta_2 = \eta_3 = 0$, it reduces to $(b^{4\kappa} + d^{4\kappa})/2 < 1$.

Finally, denote by $\mathbb{F} = (\mathcal{F}_n)$ the natural filtration of the R-BAR process $(X_k)_{k \in \mathbb{T}}$, which means that \mathcal{F}_n is the σ -algebra generated by all individuals up to the n -th generation, $\mathcal{F}_n = \sigma\{X_k, k \in \mathbb{T}_n\}$. We also introduce the sigma field $\mathcal{O} = \sigma\{\delta_k, k \in \mathbb{T}\}$ generated by the observation process. We assume that all the history of the observation process (δ_k) is known at time 0 and use the filtration $\mathbb{F}^{\mathcal{O}} = (\mathcal{F}_n^{\mathcal{O}})$ defined for all n by

$$\mathcal{F}_n^{\mathcal{O}} = \mathcal{O} \vee \sigma\{\delta_k X_k, k \in \mathbb{T}_n\} = \mathcal{O} \vee \sigma\{X_k, k \in \mathbb{T}_n^*\}.$$

Note that $\mathcal{F}_n^{\mathcal{O}}$ is a sub- σ -field of $\mathcal{O} \vee \mathcal{F}_n$.

3 Estimation

We now give some least-squares estimators of our main parameters and state our main results on their asymptotic behavior.

3.1 Estimators

We propose to make use of the standard least-squares (LS) estimator $\hat{\boldsymbol{\theta}}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n)^t$ of $\boldsymbol{\theta} = (a, b, c, d)^t$ which minimizes the following expression

$$\Delta_n(\boldsymbol{\theta}) = \frac{1}{2} \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k} (X_{2k} - a - bX_k)^2 + \delta_{2k+1} (X_{2k+1} - c - dX_k)^2.$$

Consequently, we have for all $n \geq 1$ and $i \in \{0, 1\}$

$$\hat{\boldsymbol{\theta}}_n = \mathbf{S}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\delta_{2k} X_{2k}, \delta_{2k} X_k X_{2k}, \delta_{2k+1} X_{2k+1}, \delta_{2k+1} X_k X_{2k+1})^t,$$

with $\mathbf{S}_{n-1} = \begin{pmatrix} \mathbf{S}_{n-1}^0 & 0 \\ 0 & \mathbf{S}_{n-1}^1 \end{pmatrix}$, and $\mathbf{S}_{n-1}^i = \sum_{k \in \mathbb{T}_{n-1}} \delta_{2k+i} \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}$.

We now turn to the estimation of the parameters of the conditional covariance of $(\varepsilon_2, \eta_2, \varepsilon_3, \eta_3)$. Following [21], we obtain a modified least squares estimator of $\boldsymbol{\sigma} = (\sigma_\varepsilon^2, \rho_{00}, \rho_{11}, \sigma_\eta^2)^t$ by minimizing

$$\Delta'_n(\boldsymbol{\sigma}) = \frac{1}{2} \sum_{\ell=1}^{n-1} \sum_{k \in \mathbb{G}_\ell} (\hat{\varepsilon}_{2k}^2 - \mathbb{E}[\varepsilon_{2k}^2 | \mathcal{F}_\ell^{\mathcal{O}}])^2 + (\hat{\varepsilon}_{2k+1}^2 - \mathbb{E}[\varepsilon_{2k+1}^2 | \mathcal{F}_\ell^{\mathcal{O}}])^2,$$

where for all $k \in \mathbb{G}_n$,

$$\begin{cases} \varepsilon_{2k} &= \delta_{2k}(\varepsilon_{2k} + \eta_{2k} X_k), & \hat{\varepsilon}_{2k} &= \delta_{2k}(X_{2k} - \hat{a}_n - \hat{b}_n X_k), \\ \varepsilon_{2k+1} &= \delta_{2k+1}(\varepsilon_{2k+1} + \eta_{2k+1} X_k), & \hat{\varepsilon}_{2k+1} &= \delta_{2k}(X_{2k+1} - \hat{c}_n - \hat{d}_n X_k). \end{cases}$$

Under assumptions **(H.2)** and **(H.3)**, one obtains the following estimator

$$\hat{\boldsymbol{\sigma}}_n = \mathbf{U}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\hat{\varepsilon}_{2k}^2 + \hat{\varepsilon}_{2k+1}^2, 2X_k \hat{\varepsilon}_{2k}^2, 2X_k \hat{\varepsilon}_{2k+1}^2, X_k^2 (\hat{\varepsilon}_{2k}^2 + \hat{\varepsilon}_{2k+1}^2))^t, \quad (3.1)$$

where

$$U_n = \sum_{k \in \mathbb{T}_n} \begin{pmatrix} \delta_{2k} + \delta_{2k+1} & 2\delta_{2k}X_k & 2\delta_{2k+1}X_k & (\delta_{2k} + \delta_{2k+1})X_k^2 \\ 2\delta_{2k}X_k & 4\delta_{2k}X_k^2 & 0 & 2\delta_{2k}X_k^3 \\ 2\delta_{2k+1}X_k & 0 & 4\delta_{2k+1}X_k^2 & 2\delta_{2k+1}X_k^3 \\ (\delta_{2k} + \delta_{2k+1})X_k^2 & 2\delta_{2k}X_k^3 & 2\delta_{2k+1}X_k^3 & (\delta_{2k} + \delta_{2k+1})X_k^4 \end{pmatrix}.$$

Note that if $\sigma_\eta^2 = 0$ the estimator of σ_ε^2 above corresponds to the empirical estimator already used in [15]. Similarly, the least-squares estimator of $\boldsymbol{\rho} = (\rho_\varepsilon, \rho, \rho_\eta)^t$ minimizes

$$\Delta_n''(\boldsymbol{\rho}) = \frac{1}{2} \sum_{\ell=1}^{n-1} \sum_{k \in \mathbb{G}_\ell} (\widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1} - \mathbb{E}[\epsilon_{2k} \epsilon_{2k+1} | \mathcal{F}_\ell^{\mathcal{O}}])^2,$$

and one obtains

$$\widehat{\boldsymbol{\rho}}_n = \mathbf{V}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1}, 2X_k \widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1}, X_k^2 \widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1})^t, \quad (3.2)$$

where

$$\mathbf{V}_n = \sum_{k \in \mathbb{T}_n} \delta_{2k} \delta_{2k+1} \begin{pmatrix} 1 & 2X_k & X_k^2 \\ 2X_k & 4X_k^2 & 2X_k^3 \\ X_k^2 & 2X_k^3 & X_k^4 \end{pmatrix}.$$

Note that one cannot identify ρ_{01} from ρ_{10} , hence the use of $\rho = (\rho_{01} + \rho_{10})/2$. Again if $\sigma_\eta^2 = 0$, we retrieve the empirical estimator of ρ_ε used in [15].

3.2 Main results

We now state our main results. The first one establishes the consistency of our estimators on the non-extinction set.

Theorem 3.1 *Under assumptions (H.1-5), and if $\kappa \geq 2$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \widehat{\boldsymbol{\theta}}_n = \boldsymbol{\theta} \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s.$$

and if in addition $\kappa \geq 4$ then one also has

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \widehat{\boldsymbol{\sigma}}_n = \boldsymbol{\sigma} \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s., \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \widehat{\boldsymbol{\rho}}_n = \boldsymbol{\rho} \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s.$$

The next results give convergence rates for the estimators.

Theorem 3.2 *Under assumptions (H.1-5) and if $\kappa \geq 4$, one has*

$$\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = o(n^\delta m^{-n}) \quad a.s.$$

for all $\delta > 1/2$, and the quadratic strong law

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t \mathbf{S} \boldsymbol{\Sigma}^{-1} \mathbf{S} (\widehat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) = \text{tr}(\boldsymbol{\Gamma} \boldsymbol{\Sigma}^{-1}) \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s.$$

where \mathbf{S} , $\boldsymbol{\Gamma}$ and $\boldsymbol{\Sigma}$ are 4×4 matrices defined respectively in Proposition 4.14, Lemma 5.4 and Lemma 5.5.

For all n , set

$$\boldsymbol{\sigma}_n = \mathbf{U}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\epsilon_{2k}^2 + \epsilon_{2k+1}^2, 2X_k \epsilon_{2k}^2, 2X_k \epsilon_{2k+1}^2, X_k^2 (\epsilon_{2k}^2 + \epsilon_{2k+1}^2)), \quad (3.3)$$

$$\boldsymbol{\rho}_n = \mathbf{V}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\epsilon_{2k} \epsilon_{2k+1}, 2X_k \epsilon_{2k} \epsilon_{2k+1}, X_k^2 \epsilon_{2k} \epsilon_{2k+1}). \quad (3.4)$$

Theorem 3.3 *Under assumptions (H.1-5) and if $\kappa \geq 8$, one has*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \boldsymbol{\sigma}_n &= \boldsymbol{\sigma} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_{n-1}^*|}{n} (\hat{\boldsymbol{\sigma}}_n - \boldsymbol{\sigma}_n) \\ &= \mathbf{U}^{-1} (q_0(0) + q_1(0), 2q_0(1), 2q_1(1), q_0(2) + q_1(2))^t \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned}$$

where \mathbf{U} is a 4×4 matrix defined in Proposition 4.14 and the $q_i(r)$ are scalars defined in Lemma 5.8.

Theorem 3.4 *Under assumptions (H.1-5) and if $\kappa \geq 8$, one has*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \boldsymbol{\rho}_n &= \boldsymbol{\rho} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{|\mathbb{T}_{n-1}^*|}{n} (\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}_n) &= \mathbf{V}^{-1} (q_{01}(0), 2q_{01}(1), q_{01}(2))^t \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned}$$

where \mathbf{V} is a 3×3 matrix defined in Proposition 4.14 and the $q_{01}(r)$ are scalars defined in Lemma 5.11.

We now turn to the asymptotic normality for all our estimators $\hat{\boldsymbol{\theta}}_n$, $\hat{\boldsymbol{\sigma}}_n$ and $\hat{\boldsymbol{\rho}}_n$ given the non-extinction of the underlying Galton-Watson process. Using the fact that $\mathbb{P}(\bar{\mathcal{E}}) \neq 0$ thanks to the super-criticality assumption (H.4), we define the probability $\mathbb{P}_{\bar{\mathcal{E}}}$ on (Ω, \mathcal{A}) by $\mathbb{P}_{\bar{\mathcal{E}}}(A) = \mathbb{P}(A \cap \bar{\mathcal{E}}) / \mathbb{P}(\bar{\mathcal{E}})$ for all $A \in \mathcal{A}$.

Theorem 3.5 *Under assumptions (H.1-5) and if $\kappa \geq 8$, one has*

$$|\mathbb{T}_{n-1}^*|^{1/2} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{S}^{-1} \boldsymbol{\Gamma} \mathbf{S}^{-1}) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}) \quad (3.5)$$

with \mathbf{S} defined in Proposition 4.14 and $\boldsymbol{\Gamma}$ in Lemma 5.4. If moreover $\kappa \geq 16$,

$$|\mathbb{T}_{n-1}^*|^{1/2} (\hat{\boldsymbol{\sigma}}_n - \boldsymbol{\sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{U}^{-1} \boldsymbol{\Gamma}^\sigma \mathbf{U}^{-1}) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}), \quad (3.6)$$

$$|\mathbb{T}_{n-1}^*|^{1/2} (\hat{\boldsymbol{\rho}}_n - \boldsymbol{\rho}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}^{-1} \boldsymbol{\Gamma}^\rho \mathbf{V}^{-1}) \quad \text{on } (\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}}), \quad (3.7)$$

where the matrices $\boldsymbol{\Gamma}^\sigma$ and $\boldsymbol{\Gamma}^\rho$ are defined in Eq. (6.1) and (6.2).

The proofs of these theorems are detailed in the next sections.

4 Bifurcating Markov chains and consistency

In order to investigate the convergence of our estimators, we need laws of large numbers for quantities such as $(\delta_{2k+i} X_k^q X_{2k}^r X_{2k+1}^s)_{k \in \mathbb{T}}$. To obtain them, we use the bifurcating Markov chain framework introduced by J. Guyon in [12] and adapted to Galton-Watson trees by J.-F. Delmas and L. Marsalle in [14]. Note that we cannot directly use the results in [14] because our noise sequences do not have moments of all order. Therefore, our first step is to provide a general result for bifurcating Markov chains on GW trees with only a finite number of moments. We first recall the general framework, then prove the ergodicity of the induced Markov chain and finally derive strong laws of large numbers. We conclude this section by establishing the strong consistency of our estimators.

4.1 Bifurcating Markov chain

Let \mathbb{B} be the Borel σ -field of \mathbb{R} , and \mathbb{B}^p be the Borel σ -field of \mathbb{R}^p , for $p > 1$. We add a cemetery point ∂ to \mathbb{R} , denote by $\overline{\mathbb{R}}$ the set $\mathbb{R} \cup \{\partial\}$, and by $\overline{\mathbb{B}}$ the σ -field generated by \mathbb{B} and $\{\partial\}$. This cemetery point models the state of a non-observed cell. We recall the following definitions from [14].

Definition 4.1 *We call \mathbb{T}^* -transition probability any mapping P from $\overline{\mathbb{R}} \times \overline{\mathbb{B}}^2$ onto $[0, 1]$ such that*

- $P(\cdot, A)$ is measurable for all A in $\overline{\mathbb{B}}^2$,
- $P(x, \cdot)$ is a probability measure on $(\overline{\mathbb{R}}^2, \overline{\mathbb{B}}^2)$ for all x in $\overline{\mathbb{R}}$,
- $P(\partial, \{(\partial, \partial)\}) = 1$.

For any measurable function f from $\overline{\mathbb{R}}^3$ onto \mathbb{R} , one defines the measurable function Pf from $\overline{\mathbb{R}}$ onto \mathbb{R} by

$$Pf(x) = \int f(x, y, z) P(x, dy, dz),$$

provided the integral is well defined. Let ν be a probability measure on $\overline{\mathbb{R}}$. In the sequel, ν will denote the distribution of X_1 .

Definition 4.2 *We say that $(Z_n)_{n \in \mathbb{T}}$ is a bifurcating Markov chain with initial distribution ν and \mathbb{T}^* -transition probability P , a P -BMC in short, if Z_1 has distribution ν and for all n in \mathbb{N} , and for all families of measurable bounded functions $(f_k)_{k \in \mathbb{G}_n}$ on \mathbb{R}^2 , one has*

$$\mathbb{E} \left[\prod_{k \in \mathbb{G}_n} f_k(Z_{2k}, Z_{2k+1}) \mid \sigma(Z_j, j \in \mathbb{T}_n) \right] = \prod_{k \in \mathbb{G}_n} Pf_k(Z_k).$$

As explained in [12], this means that given the first n generations \mathbb{T}_n , one builds generation \mathbb{G}_{n+1} by drawing 2^n independent couples (Z_{2k}, Z_{2k+1}) according to $P(Z_k, \cdot)$, $k \in \mathbb{G}_n$. In addition, any couple (Z_{2k}, Z_{2k+1}) depends on past generations only

through its mother Z_k . As $P(\partial, \{(\partial, \partial)\}) = 1$, ∂ is an absorbing state, and this hypothesis corresponds to the fact that a cell that is not observed cannot give birth to an observed one. We also assume that $P(x, \mathbb{R}^2)$, $P(x, \mathbb{R} \times \{\partial\})$ and $P(x, \{\partial\} \times \mathbb{R})$ do not depend on $x \in \mathbb{R}$. The P -BMC is thus said to be spatially homogeneous. Such a spatially homogeneous P -BMC with an absorbing cemetery state is called a bifurcating Markov chain on a Galton Watson tree, see [14] for details.

Now let us turn back to our observed R-BAR process. In order to use the framework of P -BMC's, we define the auxilliary process $(X_n^*)_{n \in \mathbb{T}}$ by

$$X_n^* = X_n \mathbb{1}_{\{\delta_n=1\}} + \partial \mathbb{1}_{\{\delta_n=0\}}, \quad (4.1)$$

which means that $X_n^* = X_n$ if cell n is observed, $X_n^* = \partial$ the cemetery state otherwise. It is clear from assumptions **(H.1)** and **(H.3)** that the process $(X_n^*)_{n \in \mathbb{T}}$ is a P -BMC on a GW tree with \mathbb{T}^* -transition probability given for all x in \mathbb{R} and all measurable non-negative functions f on $\overline{\mathbb{R}}^3$ by

$$\begin{aligned} Pf(x) &= p_{01} \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, (d + \eta_3)x + c + \varepsilon_3)] \\ &\quad + p_0 \mathbb{E} [f(x, (b + \eta_2)x + a + \varepsilon_2, \partial)] \\ &\quad + p_1 \mathbb{E} [f(x, \partial, (d + \eta_3)x + c + \varepsilon_3)] + (1 - p_{01} - p_0 - p_1) f(x, \partial, \partial), \end{aligned} \quad (4.2)$$

if $x \neq \partial$ and $Pf(\partial) = f(\partial, \partial, \partial)$. As explained in [12], the asymptotic behavior of the P -BMC is driven by that of the *induced* Markov chain (Y_n) defined on \mathbb{R} as follows.

- For all $n \geq 1$, define the sequence $(A_n, B_n)_{n \geq 1}$ to be i.i.d. random variables with the same distribution as $(a_{2+\zeta} + \varepsilon_{2+\zeta}, b_{2+\zeta} + \eta_{2+\zeta})$, where ζ is a Bernoulli random variable with mean $(p_{01} + p_1)/m$ independent from $(\varepsilon_2, \eta_2, \varepsilon_3, \eta_3)$.
- Then, set $Y_0 = X_1^* = X_1$ and Y_{n+1} is recursively defined by

$$Y_{n+1} = A_{n+1} + B_{n+1} Y_n. \quad (4.3)$$

The sequence $(Y_n)_{n \in \mathbb{N}}$ is clearly an \mathbb{R} -valued Markov chain with transition kernel given for all x in \mathbb{R} and A in \mathbb{B} by

$$Q(x, A) = \frac{P_0(x, A) + P_1(x, A)}{m}, \quad (4.4)$$

with $P_i(x, A) = (p_{01} + p_i) \mathbb{E} [\mathbb{1}_A((b_{2+i} + \eta_{2+i})x + a_{2+i} + \varepsilon_{2+i})]$. Note that P_0 and P_1 are sub-probability kernels on (\mathbb{R}, \mathbb{B}) , whereas Q is a proper probability kernel on (\mathbb{R}, \mathbb{B}) .

4.2 Ergodicity of the induced Markov chain

We now turn to the ergodicity for the induced Markov chain $(Y_n)_{n \in \mathbb{N}}$. We start with some preliminary results on the random variables A_1 and B_1 .

Lemma 4.3 *Under assumptions **(H.2)** and **(H.5)**, the random variables A_1 and B_1 have moments of all order up to 4γ . In addition, $\mathbb{E}[\log |B_1|] < 0$ and for all $0 < s \leq 4\kappa$, one has $\mathbb{E}[|B_1|^s] < 1$.*

Proof First, for all $0 \leq s \leq 4\gamma$, one clearly has

$$\begin{aligned}\mathbb{E}[|A_1|^s] &= \frac{p_0 + p_{01}}{m} \mathbb{E}[|a + \varepsilon_2|^s] + \frac{p_1 + p_{01}}{m} \mathbb{E}[|c + \varepsilon_3|^s], \\ \mathbb{E}[|B_1|^s] &= \frac{p_0 + p_{01}}{m} \mathbb{E}[|b + \eta_2|^s] + \frac{p_1 + p_{01}}{m} \mathbb{E}[|d + \eta_3|^s].\end{aligned}$$

Hence, under assumption **(H.2)**, it is clear that $\mathbb{E}[|A_1|^s]$ and $\mathbb{E}[|B_1|^s]$ are finite. Next, notice that the function $s \mapsto \mathbb{E}[|B_1|^s]$ is convex, that $\mathbb{E}[|B_1|^0] = 1$ and $\mathbb{E}[|B_1|^{4\kappa}] < 1$ by assumption **(H.5)**. This implies that $\mathbb{E}[|B_1|^s] < 1$ for all $0 < s \leq 4\kappa$. Last, consider $\mathbb{E}[|\log |B_1||]$: if it is finite, $\mathbb{E}[\log |B_1|]$ is the right-derivative at 0 of $s \mapsto \mathbb{E}[|B_1|^s]$, and convexity arguments with assumption **(H.5)** yield that $\mathbb{E}[\log |B_1|] < 0$; if it is infinite, the moment assumptions on B_1 gives that necessarily $\mathbb{E}[(\log |B_1|)^+] < \infty$ and $\mathbb{E}[(\log |B_1|)^-] = \infty$, so that finally $\mathbb{E}[\log |B_1|] = -\infty < 0$, as expected. \square

The next result states the existence of an invariant distribution for the Markov chain $(Y_n)_{n \in \mathbb{N}}$. It is well known as (Y_n) is a real-valued auto-regressive process with random i.i.d. coefficients satisfying Lemma 4.3, see e.g. [25, 26].

Lemma 4.4 *Under assumptions **(H.2)** and **(H.5)**, there exists a probability distribution μ on (\mathbb{R}, \mathbb{B}) which is the distribution of the convergent series $Y_\infty = \sum_{\ell=1}^{\infty} B_1 B_2 \cdots B_{\ell-1} A_\ell$ such that for all continuous bounded functions f on \mathbb{R} and all x in \mathbb{R} , one has*

$$\mathbb{E}_x[f(Y_n)] \xrightarrow{n \rightarrow \infty} \int f d\mu = \langle \mu, f \rangle.$$

We investigate the moments of the invariant distribution μ to extend the above result to polynomial functions. For all $s \geq 1$, set $\|X\|_s = (\mathbb{E}[|X|^s])^{1/s}$.

Lemma 4.5 *Under assumptions **(H.2)** and **(H.5)**, μ has moments of all order up to 4κ . In addition, for all $1 \leq s \leq 4\kappa$, all $x \in \mathbb{R}$ and all $n \in \mathbb{N}$, $(\mathbb{E}_x[|Y_n|^s])^{1/s} \leq |x| + \|A_1\|_s / (1 - \|B_1\|_s) < \infty$.*

Proof Set $1 \leq s \leq 4\kappa$. As the sequence (A_n, B_n) is i.i.d., one has

$$\mathbb{E}[|Y_\infty|^s]^{1/s} = \mathbb{E}\left[\left|\sum_{\ell=1}^{\infty} B_1 \cdots B_{\ell-1} A_\ell\right|^s\right]^{1/s} \leq \sum_{\ell=1}^{\infty} \|B_1\|_s^{\ell-1} \|A_1\|_s.$$

Since $\mathbb{E}[|B_1|^s] < 1$ and $\mathbb{E}[|A_1|^s] < \infty$ thanks to Lemma 4.3, the series converges. Now let us turn to Y_n . The recursive equation (4.3) yields

$$Y_n = Y_0 B_1 \cdots B_n + \sum_{\ell=1}^n B_n \cdots B_{\ell+1} A_\ell,$$

with the usual convention that an empty product equals 1. As the sequence (A_n, B_n) is i.i.d., Y_n also has the same distribution (under \mathbb{P}_x) as

$$x B_1 \cdots B_n + \sum_{\ell=1}^n B_1 \cdots B_{\ell-1} A_\ell, \tag{4.5}$$

so that, for $1 \leq s \leq 4\kappa$, one has

$$\mathbb{E}_x[|Y_n|^s]^{1/s} \leq |x| \|B_1\|_s^n + \sum_{\ell=1}^n \|B_1\|_s^{\ell-1} \|A_1\|_s \leq |x| + \frac{\|A_1\|_s}{1 - \|B_1\|_s},$$

hence the result. \square

Corollary 4.6 *Under assumptions (H.2) and (H.5), all polynomial functions f of degree less than 4κ are in $L^1(\mu)$: $\langle \mu, |f| \rangle = \mathbb{E}[|f(Y_\infty)|] < \infty$.*

We state a technical domination result that will be useful in the next section.

Lemma 4.7 *Under assumptions (H.2) and (H.5), for all polynomials f of degree less than $2q$ with $q \leq 2\kappa$, there exists a nonnegative polynomial g of degree less than $2q$ such that for all $n \in \mathbb{N}$ and all $x \in \mathbb{R}$ one has*

$$\left| \mathbb{E}_x[f(Y_n)] \right| \leq g(x).$$

Proof By linearity, it is sufficient to prove the result for $f(x) = x^p$ with $p \leq 2q$. For $p \geq 1$, Lemma 4.5 yields

$$\left| \mathbb{E}_x[Y_n^p] \right| \leq \left(|x| + \frac{\|A_1\|_p}{1 - \|B_1\|_p} \right)^p \leq 2^{p-1} \left(|x|^p + \frac{\|A_1\|_p^p}{(1 - \|B_1\|_p)^p} \right).$$

If p is even, we set $g(x) = 2^{p-1} (x^p + \|A_1\|_p^p / (1 - \|B_1\|_p)^p)$, and if p is odd, we set $g(x) = 2^{p-1} (x^{p+1} + 1 + \|A_1\|_p^p / (1 - \|B_1\|_p)^p)$, as for all $x \in \mathbb{R}$, $|x|^p \leq x^{p+1} + 1$. Notice that if p is odd and $p \leq 2q$, one also has $p + 1 \leq 2q$, hence the result. \square

Finally, we prove the geometric ergodicity of (Y_n) for polynomial functions.

Lemma 4.8 *Under assumptions (H.1-2) and (H.5), for all polynomial functions f of degree less than $2q$ with $q \leq 2\kappa$, there exist a nonnegative polynomial function g of degree less than $2q$ and a positive constant c such that for all $n \in \mathbb{N}$ and all $x \in \mathbb{R}$, one has*

$$\left| \mathbb{E}_x[f(Y_n)] - \langle \mu, f \rangle \right| \leq g(x) \|B_1\|_{4\kappa}^n, \quad \text{and} \quad \left| \mathbb{E}_\nu[f(Y_n)] - \langle \mu, f \rangle \right| \leq c \|B_1\|_{4\kappa}^n.$$

Proof Without loss of generality, it is sufficient to prove the result for polynomials f of the form x^p with $1 \leq p \leq 2q$. Hölder inequality yields

$$\begin{aligned} \left| \mathbb{E}_x[f(Y_n)] - \langle \mu, f \rangle \right| &= \left| \mathbb{E}_x[Y_n^p - Y_\infty^p] \right| = \left| \mathbb{E}_x[(Y_n - Y_\infty) \sum_{s=0}^{p-1} Y_n^s Y_\infty^{p-1-s}] \right| \\ &\leq \left(\mathbb{E}_x[|Y_n - Y_\infty|^p] \right)^{\frac{1}{p}} \sum_{s=0}^{p-1} \left(\mathbb{E}_x[|Y_n^s Y_\infty^{p-1-s}|^{\frac{p}{p-1}}] \right)^{\frac{p-1}{p}}. \end{aligned}$$

We are going to study the two terms above separately. For the first term, Eq. (4.5) and the definition of Y_∞ yield

$$\begin{aligned} \left(\mathbb{E}_x[|Y_n - Y_\infty|^p]\right)^{1/p} &= \left(\mathbb{E}[|xB_1 \cdots B_n - \sum_{\ell=n+1}^{\infty} B_1 \cdots B_{\ell-1} A_\ell|^p]\right)^{1/p} \\ &\leq |x| \|B_1\|_p^n + \|A_1\|_p \frac{\|B_1\|_p^n}{1 - \|B_1\|_p} \\ &\leq \left(|x| + \frac{\|A_1\|_p}{1 - \|B_1\|_p}\right) \|B_1\|_{4\kappa}^n, \end{aligned}$$

by Lemma 4.3 as $p \leq 4\kappa$ by assumption. We now turn to the second term. Hölder inequality with parameters $(p-1)/s$ and $(p-1)/(p-1-s)$ yields

$$\begin{aligned} \left(\mathbb{E}_x[|Y_n^s Y_\infty^{p-1-s}|^{\frac{p}{p-1}}]\right) &\leq \left(\mathbb{E}_x[|Y_n|^p]\right)^{s/(p-1)} \left(\mathbb{E}[|Y_\infty|^p]\right)^{(p-1-s)/(p-1)} \\ &\leq \left(|x| + \frac{\|A_1\|_p}{1 - \|B_1\|_p}\right)^s \|Y_\infty\|_p^{p-1-s}, \end{aligned}$$

this last majoration coming from Lemma 4.5. Finally, one obtains

$$\begin{aligned} \left|\mathbb{E}_x[Y_n^p - Y_\infty^p]\right| &\leq \|B_1\|_{4\kappa}^n \sum_{s=0}^{p-1} \left(|x| + \frac{\|A_1\|_p}{1 - \|B_1\|_p}\right)^{s+1} \|Y_\infty\|_p^{p-1-s} \\ &\leq \|B_1\|_{4\kappa}^n g(x), \end{aligned}$$

where g is a polynomial function of degree at most $2q$ by a similar argument as in the previous proof. Integrating this bound with respect to the initial law ν and using **(H.1)** gives the second result. \square

4.3 Laws of large numbers for the P -BMC

We now want to prove laws of large numbers for a family of functionals of the P -BMC (X_n^*). We are interested in polynomial functions on $\overline{\mathbb{R}}$ and $\overline{\mathbb{R}}^3$ multiplied by indicators. Precisely, for all $q \geq 1$, let F_q and G_q be the vector spaces generated by the following classes of functions from $\overline{\mathbb{R}}^3$ onto \mathbb{R} and from $\overline{\mathbb{R}}$ onto \mathbb{R} respectively,

$$F_q = \text{vect}\{x^\alpha y^\beta \mathbb{1}_{\mathbb{R}}(y), x^\alpha z^\tau \mathbb{1}_{\mathbb{R}}(z), x^\alpha y^\beta z^\tau \mathbb{1}_{\mathbb{R}^2}(y, z), \quad 0 \leq \alpha + \beta + \tau \leq q\},$$

$$G_q = \text{vect}\{x^\alpha \mathbb{1}_{\mathbb{R}}(x), \quad 0 \leq \alpha \leq q\},$$

where α, β, τ are integers. We first establish some technical results.

Lemma 4.9 *Let $f \in F_q$ and $h \in G_q$. Under assumption **(H.2)**,*

- (i) *if $q \leq 4\gamma$, then $f \in L^1(P)$ and $Pf \in G_q$,*
- (ii) *if $q \leq 4\gamma$, then $h \in L^1(P_0) \cap L^1(P_1) \cap L^1(Q)$ and P_0h, P_1h and $Qh \in G_q$,*
- (iii) *if $q \leq 2\gamma$, then $h \otimes h \in L^1(P)$ and $P(h \otimes h) \in G_{2q}$.*

Proof Take $q \leq 4\gamma$ and remark that $Pf(\partial) = 0$ for any $f \in F_q$, so that $Pf(x) = Pf(x)\mathbb{1}_{\mathbb{R}}(x)$ for all $x \in \overline{\mathbb{R}}$. Next, set $f(x, y, z) = x^\alpha y^\beta z^\tau \mathbb{1}_{\mathbb{R}^2}(y, z)$, $f_0 = x^\alpha y^\beta \mathbb{1}_{\mathbb{R}}(y)$ and $f_1 = x^\alpha z^\beta \mathbb{1}_{\mathbb{R}}(z)$ in F_q . Eq. (4.2) yields, for $i \in \{0, 1\}$,

$$\begin{aligned} P|f|(x) &= p_{01}|x|^\alpha \mathbb{E}[|(b + \eta_2)x + a + \varepsilon_2|^\beta |(d + \eta_3)x + c + \varepsilon_3|^\tau], \\ P|f_i|(x) &= (p_{01} + p_i)|x|^\alpha \mathbb{E}[|(b_{2+i} + \eta_{2+i})x + a_{2+i} + \varepsilon_{2+i}|^\beta]. \end{aligned}$$

Assumption **(H.2)** entails that the 4γ moments of $((b + \eta_2)x + a + \varepsilon_2)$ and $((d + \eta_3)x + c + \varepsilon_3)$ are finite, which gives the finiteness of all the above expectations, since $\beta + \tau \leq q$. The results of integrability for (i) is thus proved by linearity. In addition, one has, for $i \in \{0, 1\}$,

$$\begin{aligned} Pf(x) &= p_{01} \sum_{r=0}^{\beta} \sum_{s=0}^{\tau} C_\beta^r C_\tau^s \mathbb{E}[(b + \eta_2)^r (a + \varepsilon_2)^{\beta-r} (d + \eta_3)^s (c + \varepsilon_3)^{\tau-s}] x^{r+s+\alpha}, \\ Pf_i(x) &= (p_{01} + p_i) \sum_{r=0}^{\beta} C_\beta^r \mathbb{E}[(b_{2+i} + \eta_{2+i})^r (a_{2+i} + \varepsilon_{2+i})^{\beta-r}] x^{r+\alpha}, \end{aligned}$$

so that Pf , Pf_1 and Pf_2 in G_q , since $\alpha + \beta + \tau \leq q$, and (i) is obtained by linearity. Now set $h(x) = x^\alpha \mathbb{1}_{\mathbb{R}}(x)$ in G_q . One has

$$P|h|(x) = (p_{01} + p_i) \mathbb{E}[|(b_{2+i} + \eta_{2+i})x + a_{2+i} + \varepsilon_{2+i}|^\alpha],$$

which is similar to Pf_i so that the same arguments as above imply (ii) . Finally, set $p \leq 2\gamma$, $h_0(x) = x^\beta \mathbb{1}_{\mathbb{R}}(x)$, $h_1(x) = x^\tau \mathbb{1}_{\mathbb{R}}(x)$ in G_p and $l(y, z) = h_0 \otimes h_1(y, z) = h_0(y)h_1(z)$. One has

$$P|l|(x) = p_{01} \mathbb{E}[|(b + \eta_2)x + a + \varepsilon_2|^\beta |(d + \eta_3)x + c + \varepsilon_3|^\tau],$$

which is similar to Pf with $\beta + \tau \leq 2p$ so that (iii) also holds. \square

We are now ready to prove the main result of this section.

Theorem 4.10 *Under assumptions **(H.1-5)**, for all function $f \in F_\kappa$, one has the following law of large numbers*

$$\lim_{n \rightarrow \infty} \frac{1}{m^n} \sum_{k \in \mathbb{G}_n^*} f(X_k^*, X_{2k}^*, X_{2k+1}^*) = \langle \mu, Pf \rangle W \quad a.s.$$

Proof: This result is similar to Theorem 11 of [12] and Theorem 3.1 of [14]. The proof follows essentially the same lines and is thus shortened here, the main difference being that the class of functions F_κ does not satisfy assumptions (i)-(vi) from [12, 14] mainly because F_κ is not stable by multiplication and $(\varepsilon_2, \eta_2, \varepsilon_3, \eta_3)$ do not have moments of all order.

For all f in F_κ , Pf is well-defined from $\overline{\mathbb{R}}$ onto \mathbb{R} thanks to Lemma 4.9 as $\kappa \leq \gamma$. As $Pf(\partial) = 0$, by a slight abuse of notation we will also denote Pf its restriction

to \mathbb{R} . Thus, Pf is μ -integrable by Lemma 4.5. One has

$$\begin{aligned} & m^{-n} \sum_{k \in \mathbb{G}_n^*} f(X_k^*, X_{2k}^*, X_{2k+1}^*) - \langle \mu, Pf \rangle W \\ &= m^{-n} \sum_{k \in \mathbb{G}_n^*} (f(X_k^*, X_{2k}^*, X_{2k+1}^*) - \langle \mu, Pf \rangle) + \langle \mu, Pf \rangle \left(\frac{|\mathbb{G}_n^*|}{m^n} - W \right). \end{aligned}$$

By Eq. (2.3) the second term converges to 0 a.s. as n tends to infinity. In order to prove the a.s. convergence of the first term, as in [12, 14], it is sufficient to prove that

$$\sum_{n \geq 0} m^{-2n} \mathbb{E} \left[\left(\sum_{k \in \mathbb{G}_n^*} g(X_k^*, X_{2k}^*, X_{2k+1}^*) \right)^2 \right] < \infty, \quad (4.6)$$

with $g = f - \langle \mu, Pf \rangle \in F_\kappa$. Thanks to Lemma 4.9, $Pg \in G_\kappa$, and as $g^2 \in F_{2\kappa}$, one also has $Pg^2 \in G_{2\kappa}$. The expectation inside the sum decomposes as

$$\begin{aligned} & \mathbb{E} \left[\left(\sum_{k \in \mathbb{G}_n^*} g(X_k^*, X_{2k}^*, X_{2k+1}^*) \right)^2 \right] \\ &= \mathbb{E} \left[\left(\sum_{k \in \mathbb{G}_n^*} Pg(X_k^*) \right)^2 \right] + \mathbb{E} \left[\sum_{k \in \mathbb{G}_n^*} (Pg^2 - (Pg)^2)(X_k^*) \right] = C_n + D_n. \end{aligned}$$

We study the two terms C_n and D_n separately. Let us first prove that $\sum_{n \geq 0} m^{-2n} D_n < \infty$. We can rewrite $D_n = \mathbb{E} \left[\sum_{k \in \mathbb{G}_n^*} h(X_k^*) \right]$ with $h = Pg^2 - (Pg)^2$. As seen above, $h \in G_{2\kappa}$ and therefore h is μ -integrable thanks to Lemma 4.5. To investigate the limit of $\sum m^{-2n} D_n$, we prove that $m^{-n} D_n$ has a finite limit. More precisely, one has

$$\begin{aligned} & \left\| m^{-n} \sum_{k \in \mathbb{G}_n^*} h(X_k^*) - \langle \mu, h \rangle W \right\|_2 \\ &= \left\| m^{-n} \sum_{k \in \mathbb{G}_n^*} (h(X_k^*) - \langle \mu, h \rangle) + \langle \mu, h \rangle (m^{-n} |\mathbb{G}_n^*| - W) \right\|_2 \\ &\leq \left\| m^{-n} \sum_{k \in \mathbb{G}_n^*} (h(X_k^*) - \langle \mu, h \rangle) \right\|_2 + |\langle \mu, h \rangle| \left\| m^{-n} |\mathbb{G}_n^*| - W \right\|_2. \end{aligned}$$

The second term converges to zero. For the first term, again let $l = h - \langle \mu, h \rangle \in G_{2\kappa}$ and $\langle \mu, l \rangle = 0$, and by [14, Eq. (15) p 2504], one has

$$\begin{aligned} & \left\| m^{-n} \sum_{k \in \mathbb{G}_n^*} (h(X_k^*) - \langle \mu, h \rangle) \right\|_2^2 \quad (4.7) \\ &= m^{-n} \mathbb{E}_\nu [l^2(Y_n)] + 2m^{-2} \sum_{\ell=0}^{n-1} m^{-\ell} \langle \nu, Q^\ell P(Q^{n-\ell-1} l \otimes Q^{n-\ell-1} l) \rangle. \end{aligned}$$

Concerning the first term in Eq. (4.7), as $l^2 \in G_{4\kappa}$, by Lemma 4.8 one obtains $\lim_{n \rightarrow \infty} \mathbb{E}_\nu [l^2(Y_n)] = \langle \mu, l^2 \rangle$ and $m^{-n} \mathbb{E} [l^2(Y_n)]$ converges to 0 a.s. Concerning the second term in Eq. (4.7), Lemma 4.8 yields $\lim_{n \rightarrow \infty} Q^{n-\ell-1} l(x) = \lim_{n \rightarrow \infty} \mathbb{E}_x [l(Y_{n-\ell-1})] = \langle \mu, l \rangle = 0$ and by Lemma 4.7, $Q^{n-\ell-1} l$ is dominated by some $\phi \in G_{2\kappa}$. Moreover, using Lemma 4.9, $\phi \otimes \phi$ belongs to $F_{4\kappa}$, it is P -integrable and $P(\phi \otimes \phi)$ belongs to $G_{4\kappa}$.

By Lemma 4.7, $Q^\ell P(\phi \otimes \phi)$ is dominated by some $\psi \in G_{4\kappa}$, which is ν -integrable by assumption **(H.2)**. Lebesgue dominated convergence theorem thus yields

$$\lim_{n \rightarrow \infty} \langle \nu, Q^\ell P(Q^{n-\ell-1}l \otimes Q^{n-\ell-1}l) \rangle = 0,$$

and $|\langle \nu, Q^\ell P(Q^{n-\ell-1}l \otimes Q^{n-\ell-1}l) \rangle| \leq \langle \nu, \psi \rangle$. This upper bound allows us to deal with the limit of the second term of Eq. (4.7). Under assumption **(H.4)**, $\sum_{\ell=0}^{n-1} m^{-\ell}$ converges and for $\epsilon > 0$, it exists ℓ_ϵ such that $\sum_{\ell=\ell_\epsilon}^{n-1} m^{-\ell} \langle \nu, \psi \rangle \leq \epsilon$. Finally, for $n > \ell_\epsilon$, we have

$$\begin{aligned} & \left| \sum_{\ell=0}^{n-1} m^{-\ell} \langle \nu, Q^\ell P(Q^{n-\ell-1}l \otimes Q^{n-\ell-1}l) \rangle \right| \\ & \leq \sum_{\ell=0}^{\ell_\epsilon-1} m^{-\ell} |\langle \nu, Q^\ell P(Q^{n-\ell-1}l \otimes Q^{n-\ell-1}l) \rangle| + \epsilon, \end{aligned}$$

All the terms of the left sum converge to 0 with n , which finally proves the L_2 -convergence of $m^{-n} \sum_{k \in \mathbb{G}_n^*} h(X_k^*)$ to $\langle \mu, h \rangle W$. It implies the convergence of the expectation $m^{-n} D_n$ to $\langle \mu, h \rangle \mathbb{E}[W]$ (recall that W is square integrable). Therefore, one obtains $\sum_{n \geq 0} m^{-2n} D_n < \infty$ because $m > 1$.

Let us now prove that $\sum_{n \geq 0} m^{-2n} C_n < \infty$. Recall that $g \in F_\kappa$, $\langle \mu, Pg \rangle = 0$ and following [14, Eq. (15) p 2504], we have

$$\begin{aligned} \frac{C_n}{m^{2n}} &= \left\| \frac{1}{m^n} \sum_{k \in \mathbb{G}_n^*} Pg(X_k^*) \right\|_2^2 \\ &= \frac{1}{m^n} \mathbb{E}_\nu[(Pg)^2(Y_n)] + \frac{2}{m^2} \sum_{\ell=0}^{n-1} \frac{\langle \nu, Q^\ell P(Q^{n-\ell-1}(Pg) \otimes Q^{n-\ell-1}(Pg)) \rangle}{m^\ell}. \end{aligned} \quad (4.8)$$

The proof of the convergence of the first term of Eq. (4.8) is the same as that of $\mathbb{E}_\nu[l^2(Y_q)]$, and $\sum_{n \geq 0} m^{-n} \mathbb{E}[(Pg)^2(Y_n)]$ converges. For the second term, setting $p = n - \ell - 1$, we can rewrite

$$\begin{aligned} & \sum_{n \geq 0} \sum_{\ell=0}^{n-1} m^{-\ell} \langle \nu, Q^\ell P(Q^{n-\ell-1}(Pg) \otimes Q^{n-\ell-1}(Pg)) \rangle \\ &= \sum_{\ell \geq 0} m^{-\ell} \left\langle \nu, Q^\ell P \left(\sum_{p \geq 0} Q^p(Pg) \otimes Q^p(Pg) \right) \right\rangle. \end{aligned}$$

By Lemma 4.8, there exists $\varphi \in G_{\kappa+1}$, such that $|\mathbb{E}_x[(Pg)(Y_p)]| = |Q^p(Pg)(x)| \leq \varphi(x) \|B_1\|_{4\kappa}^p$ and therefore one has

$$\left| \sum_{p \geq 0} (Q^p(Pg) \otimes Q^p(Pg)) \right| \leq (\varphi \otimes \varphi) \sum_{p \geq 0} \|B_1\|_{4\kappa}^p.$$

By assumption **(H.5)**, the series converges and there remains to study the asymptotic behavior of $\sum_{\ell \geq 0} m^{-\ell} \langle \nu, Q^\ell P(\varphi \otimes \varphi) \rangle$. For this, let us remark that $\langle \nu, Q^\ell P(\varphi \otimes \varphi) \rangle = \mathbb{E}_\nu[P(\varphi \otimes \varphi)(\tilde{Y}_\ell)]$ with $P(\varphi \otimes \varphi) \in G_{2\kappa+2}$. By Lemma 4.8, $\lim_{\ell \rightarrow \infty} \mathbb{E}_\nu[P(\varphi \otimes \varphi)(\tilde{Y}_\ell)]$ is finite and the series converges because $m > 1$. We have thus proved that Eq. (4.6) holds, and hence the almost sure convergence of the series $m^{-n} \sum_{\ell \in \mathbb{G}_n^*} f(X_\ell^*, X_{2\ell}^*, X_{2\ell+1}^*)$ to $\langle \mu, Pf \rangle W$. \square

4.4 Laws of large numbers for the R-BAR process

Let us now turn back to our R-BAR process and see how the law of large numbers given by Theorem 4.10 applies to our process.

Proposition 4.11 *Under assumptions (H.1-5), for all integers $0 \leq q \leq \kappa$, and all $i \in \{0, 1\}$, the following laws of large numbers hold*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \delta_{2k+i} X_k^q &= \ell_i(q) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \delta_{2k} \delta_{2k+1} X_k^q &= \ell_{01}(q) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned}$$

with $\ell_i(q) = (p_{01} + p_i) \mathbb{E}[Y_\infty^q]$ and $\ell_{01}(q) = p_{01} \mathbb{E}[Y_\infty^q]$.

Proof Set $q \leq \kappa$. We apply Theorem 4.10 to the function $f_0(x, y, z) = x^q \mathbb{1}_{\mathbb{R}}(y)$ if $i = 0$ and $f_1(x, y, z) = x^q \mathbb{1}_{\mathbb{R}}(z)$ if $i = 1$ for the first limit, and $f_{01}(x, y, z) = x^q \mathbb{1}_{\mathbb{R}^2}(y, z)$ for the second limit. The functions f_0 , f_1 and f_{01} clearly belong to F_κ , and moreover $Pf_i(x) = (p_{01} + p_i)x^q$, $Pf_{01}(x) = p_{01}x^q$. Finally, notice that $\langle \mu, x^q \rangle = \mathbb{E}[Y_\infty^q]$. Theorem 4.10 thus yields a.s.

$$\lim_{n \rightarrow \infty} \frac{1}{m^n} \sum_{k \in \mathbb{G}_n^*} \delta_{2k+i} X_k^q = \ell_i(q) W, \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{m^n} \sum_{k \in \mathbb{G}_n^*} \delta_{2k} \delta_{2k+1} X_k^q = \ell_{01}(q) W.$$

Now, one has, for instance

$$\lim_{n \rightarrow \infty} \frac{1}{m^n} \sum_{k \in \mathbb{T}_n^*} \delta_{2k+i} X_k^q = \sum_{\ell=0}^n \frac{1}{m^{n-\ell}} \left(\frac{1}{m^\ell} \sum_{k \in \mathbb{G}_\ell^*} \delta_{2k+i} X_k^q \right).$$

The sum above converges to $\ell_i(q) W m / (m-1)$ thanks to [13, Lemma A.3] and we conclude using Eq. (2.4). \square

Proposition 4.12 *Under assumptions (H.1-5), for all integers $0 \leq q \leq \kappa - 1$, and all $i \in \{0, 1\}$, one has the following almost sure convergences*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k+i} X_k^q X_{2k+i} &= (p_{01} + p_i) (a_{2+i} \mathbb{E}[Y_\infty^q] + b_{2+i} \mathbb{E}[Y_\infty^{q+1}]) \mathbb{1}_{\bar{\mathcal{E}}}, \\ \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k} \delta_{2k+1} X_k^q X_{2k+i} &= p_{01} (a_{2+i} \mathbb{E}[Y_\infty^q] + b_{2+i} \mathbb{E}[Y_\infty^{q+1}]) \mathbb{1}_{\bar{\mathcal{E}}}, \end{aligned}$$

and if $\kappa \geq 2$, for all integers $0 \leq q \leq \kappa - 2$, one has a.s.

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k+i} X_k^q X_{2k+i}^2 &= (p_{01} + p_i) \\ &\quad \times ((a_{2+i}^2 + \sigma_\varepsilon^2) \mathbb{E}[Y_\infty^q] + 2(a_{2+i} b_{2+i} + \rho_{ii}) \mathbb{E}[Y_\infty^{q+1}] + (b_{2+i}^2 + \sigma_\eta^2) \mathbb{E}[Y_\infty^{q+2}]) \mathbb{1}_{\bar{\mathcal{E}}}, \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k} \delta_{2k+1} X_k^q X_{2k} X_{2k+1} &= \\ &= p_{01} ((ac + \rho_\varepsilon) \mathbb{E}[Y_\infty^q] + (ad + bc + 2\rho) \mathbb{E}[Y_\infty^{q+1}] + (bd + \rho_\eta) \mathbb{E}[Y_\infty^{q+2}]) \mathbb{1}_{\bar{\mathcal{E}}}. \end{aligned}$$

Proof The proof follows the same lines as that of Proposition 4.11. \square

We end this section by computing the moments of the invariant law μ .

Lemma 4.13 *Under assumptions (H.2) and (H.5), one has*

$$\mathbb{E}[Y_\infty] = \frac{\mathbb{E}[A_1]}{1 - \mathbb{E}[B_1]}, \quad \mathbb{E}[Y_\infty^2] = \frac{\mathbb{E}[A_1^2] + 2\mathbb{E}[A_1 B_1]\mathbb{E}[Y_\infty]}{1 - \mathbb{E}[B_1^2]},$$

and more generally, the moments of Y_∞ can be calculated recursively for all $1 \leq q \leq 4\kappa$ thanks to the relation $\mathbb{E}[Y_\infty^q] = \sum_{s=0}^q C_q^s \mathbb{E}[A_1^{q-s} B_1^s] \mathbb{E}[Y_\infty^s]$.

Proof As Y_∞ is the stationary solution of equation $Y_n = A_n + B_n Y_{n-1}$, Y_∞ has the same law as $A_0 + B_0 Y_\infty$ where (A_0, B_0) is a copy of (A_1, B_1) independent from the sequence $(A_n, B_n)_{n \geq 1}$. Hence, one has $\mathbb{E}[Y_\infty] = \mathbb{E}[A_0 + B_0 Y_\infty] = \mathbb{E}[A_1] + \mathbb{E}[B_1] \mathbb{E}[Y_\infty]$. Similarly, one has

$$\mathbb{E}[Y_\infty^2] = \mathbb{E}[(A_0 + B_0 Y_\infty)^2] = \mathbb{E}[A_1^2] + 2\mathbb{E}[A_1 B_1] \mathbb{E}[Y_\infty] + \mathbb{E}[B_1^2] \mathbb{E}[Y_\infty^2].$$

The general formula is obtained in the same way by developing the relation $\mathbb{E}[Y_\infty^q] = \mathbb{E}[(A_0 + B_0 Y_\infty)^q]$. \square

Note that one can easily compute the moments of A_1 and B_1 from their definition. In particular, one has

$$\begin{aligned} \mathbb{E}[Y_\infty] &= \frac{am_0 + cm_1}{1 - bm_0 - dm_1}, \\ \mathbb{E}[Y_\infty^2] &= \frac{a^2 m_0 + c^2 m_1 + \sigma_\varepsilon^2 + 2((ab + \rho_{00})m_0 + (cd + \rho_{11})m_1)\mathbb{E}[Y_\infty]}{1 - (b^2 m_0 + d^2 m_1 + \sigma_\eta^2)}, \end{aligned}$$

with $m_0 = (p_{01} + p_0)/m$ and $m_1 = (p_{01} + p_1)/m$.

4.5 Consistency of the estimators

We are now able to prove the consistency of our estimators. We start with the computation of the limits of the normalizing matrices \mathbf{S}_n , \mathbf{U}_n and \mathbf{V}_n , which is a direct consequence of Proposition 4.11

Proposition 4.14 *Under assumptions (H.1-5), and if $\kappa \geq 2$, for $i \in \{0, 1\}$, one has*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathbf{S}_n^i}{|\mathbb{T}_n^*|} &= \mathbf{S}^i \mathbb{1}_{\bar{\mathcal{E}}} = (p_{01} + p_i) \begin{pmatrix} 1 & \mathbb{E}[Y_\infty] \\ \mathbb{E}[Y_\infty] & \mathbb{E}[Y_\infty^2] \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathbf{S}_n}{|\mathbb{T}_n^*|} &= \mathbf{S} \mathbb{1}_{\bar{\mathcal{E}}} = \begin{pmatrix} \mathbf{S}^0 & 0 \\ 0 & \mathbf{S}^1 \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned}$$

If in addition $\kappa \geq 4$, the following convergences hold

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathbf{U}_n}{|\mathbb{T}_n^*|} &= \mathbf{U} \mathbb{1}_{\bar{\mathcal{E}}} \\ &= m \begin{pmatrix} 1 & 2m_0 \mathbb{E}[Y_\infty] & 2m_1 \mathbb{E}[Y_\infty] & \mathbb{E}[Y_\infty^2] \\ 2m_0 \mathbb{E}[Y_\infty] & 4m_0 \mathbb{E}[Y_\infty^2] & 0 & 2m_0 \mathbb{E}[Y_\infty^3] \\ 2m_1 \mathbb{E}[Y_\infty] & 0 & 4m_1 \mathbb{E}[Y_\infty^2] & 2m_1 \mathbb{E}[Y_\infty^3] \\ \mathbb{E}[Y_\infty^2] & 2m_0 \mathbb{E}[Y_\infty^3] & 2m_1 \mathbb{E}[Y_\infty^3] & \mathbb{E}[Y_\infty^4] \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s. \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathbf{V}_n}{|\mathbb{T}_n^*|} = \mathbf{V} \mathbb{1}_{\bar{\mathcal{E}}} = p_{01} \begin{pmatrix} 1 & 2\mathbb{E}[Y_\infty] & \mathbb{E}[Y_\infty^2] \\ 2\mathbb{E}[Y_\infty] & 4\mathbb{E}[Y_\infty^2] & 2\mathbb{E}[Y_\infty^3] \\ \mathbb{E}[Y_\infty^2] & 2\mathbb{E}[Y_\infty^3] & \mathbb{E}[Y_\infty^4] \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Besides, the matrices \mathbf{S}^i , \mathbf{U} and \mathbf{V} are assumed to be invertible.

We now turn to the consistency of our main estimators.

Proof of Theorem 3.1 As regards our main estimator $\hat{\boldsymbol{\theta}}_n$, a direct application of Proposition 4.12 yields

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \mathbf{S}_{n-1} \hat{\boldsymbol{\theta}}_n = m \begin{pmatrix} m_0(a + b\mathbb{E}[Y_\infty]) \\ m_0(a\mathbb{E}[Y_\infty] + b\mathbb{E}[Y_\infty^2]) \\ m_1(c + d\mathbb{E}[Y_\infty]) \\ m_1(c\mathbb{E}[Y_\infty] + d\mathbb{E}[Y_\infty^2]) \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} = \mathbf{S}\boldsymbol{\theta} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

and the result follows from Proposition 4.14 and the definition of $\hat{\boldsymbol{\theta}}_n$. The consistency of $\hat{\boldsymbol{\sigma}}_n$ and $\hat{\boldsymbol{\rho}}_n$ is more complicated as their definition involves the $\hat{\epsilon}_k$. We give a detailed proof of the convergence of $|\mathbb{T}_{n-1}^*|^{-1} \sum \hat{\epsilon}_{2k}^2$, the other terms in $\mathbf{U}_{n-1} \hat{\boldsymbol{\sigma}}_n$ and $\mathbf{V}_{n-1} \hat{\boldsymbol{\rho}}_n$ being treated similarly. For $k \in \mathbb{G}_n$, one has

$$\begin{aligned} \hat{\epsilon}_{2k}^2 &= \delta_{2k}(X_{2k} - \hat{a}_n - \hat{b}_n X_k)^2 \\ &= \delta_{2k}(\hat{a}_n^2 + 2\hat{a}_n \hat{b}_n X_k + \hat{b}_n^2 X_k^2 - 2\hat{a}_n X_{2k} - 2\hat{b}_n X_k X_{2k} + X_{2k}^2). \end{aligned}$$

Hence, one has

$$\begin{aligned} \sum_{k \in \mathbb{T}_{n-1}^*} \hat{\epsilon}_{2k}^2 &= \sum_{\ell=1}^{n-1} \hat{a}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{2k} + 2 \sum_{\ell=1}^{n-1} \hat{a}_\ell \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{2k} X_k + \sum_{\ell=1}^{n-1} \hat{b}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{2k} X_k^2 \\ &\quad - 2 \sum_{\ell=1}^{n-1} \hat{a}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{2k} X_{2k} - 2 \sum_{\ell=1}^{n-1} \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{2k} X_k X_{2k} + \sum_{k \in \mathbb{T}_{n-1}^*} \delta_{2k} X_{2k}^2. \end{aligned} \quad (4.9)$$

The limit of the last term is given by Proposition 4.12. Let us study the first term. One has

$$\frac{1}{m^{n-1}} \sum_{\ell=1}^{n-1} \hat{a}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{2k} = \sum_{\ell=1}^{n-1} \hat{a}_\ell^2 \frac{m^\ell}{m^{n-1}} \frac{1}{m^\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{2k}.$$

We apply Lemma A.3 of [13] to the sequence above. On the one hand,

$$\lim_{\ell \rightarrow \infty} \hat{a}_\ell^2 \frac{1}{m^\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{2k} = a^2(p_{01} + p_0)W \quad a.s.$$

thanks to the previous result on the consistency of $\hat{\boldsymbol{\theta}}_n$ and Theorem 4.10. On the other hand, the series $\sum m^{-n}$ converges to $m/(m-1)$ under assumption **(H.4)**. Therefore, Lemma A.3 of [13] yields

$$\lim_{n \rightarrow \infty} \frac{1}{m^{n-1}} \sum_{\ell=1}^{n-1} \hat{a}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{2k} = \frac{m}{m-1} a^2(p_{01} + p_0)W \quad a.s.$$

and Eq. (2.4) finally yields

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{\ell=1}^{n-1} \widehat{a}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{2k} = a^2(p_{01} + p_0) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

Note that the limit above is just the limit of \widehat{a}_ℓ^2 multiplied by the limit of $|\mathbb{T}_{n-1}^*|^{-1} \sum \delta_{2k}$. The other terms in Eq. (4.9) are dealt with similarly using the results of Proposition 4.12. Finally, one obtains the almost sure convergences

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \mathbf{U}_{n-1} \widehat{\boldsymbol{\sigma}}_n &= \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \widehat{\epsilon}_{2k}^2 + \widehat{\epsilon}_{2k+1}^2 \\ 2X_k \widehat{\epsilon}_{2k}^2 \\ 2X_k \widehat{\epsilon}_{2k+1}^2 \\ X_k^2 (\widehat{\epsilon}_{2k}^2 + \widehat{\epsilon}_{2k+1}^2) \end{pmatrix} = \mathbf{U} \boldsymbol{\sigma} \mathbb{1}_{\overline{\mathcal{E}}}, \\ \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \mathbf{V}_{n-1} \widehat{\boldsymbol{\rho}}_n &= \lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1} \\ 2X_k \widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1} \\ X_k^2 \widehat{\epsilon}_{2k} \widehat{\epsilon}_{2k+1} \end{pmatrix} = \mathbf{V} \boldsymbol{\rho} \mathbb{1}_{\overline{\mathcal{E}}}, \end{aligned}$$

hence the result using Proposition 4.14. \square

5 Martingales and convergence rate

The aim of this section is to obtain sharper convergence results for our estimators, namely rates of convergence. The P -BMC approach does not allow this, therefore we now use martingale theory instead, as in [13, 15]. However, we cannot directly apply the results therein mainly because our noise sequence ($\epsilon_k = \varepsilon_k + \eta_k X_{[k/2]}$) now contains the BAR process (X_k) and thus does not satisfy the assumptions of [13, 15].

5.1 Martingales on binary trees

We start with a general result of convergence for martingales on a Galton Watson binary tree, that we will make repeatedly use of in the following sections. Special cases of this result have already been proved and used in [13] and [15]. Note that in this binary tree context, we cannot make use of the standard asymptotic theory for vector martingales (see e.g. [22]) because the size of data is roughly multiplied by m at each generation.

Theorem 5.1 *Let (\mathbf{M}_n) be a p -dimensional $\mathbb{F}^{\mathcal{O}}$ -martingale on the GW-binary tree \mathbb{T}^* : $\mathbf{M}_n = \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell^*} \mathbf{W}_k$, with $\mathbf{W}_k = (w_k^1, w_k^2, \dots, w_k^p)^t$. We make the following assumptions*

(A.1) *For all n , \mathbf{M}_n is square integrable.*

Let $\langle \mathbf{M} \rangle_n = \sum_{\ell=0}^{n-1} \boldsymbol{\Gamma}_\ell$ be the increasing process of (\mathbf{M}_n) , with

$$\boldsymbol{\Gamma}_n = \mathbb{E}[\Delta \mathbf{M}_{n+1} \Delta \mathbf{M}_{n+1}^t \mid \mathcal{F}_n^{\mathcal{O}}].$$

(A.2) *On $\overline{\mathcal{E}}$, $|\mathbb{T}_{n-1}^*|^{-1} \langle \mathbf{M} \rangle_n$ converges almost surely to a positive semidefinite matrix $\boldsymbol{\Gamma}$.*

(A.3) The $p \times p$ $\mathbb{F}^\mathcal{O}$ -matrix martingale (\mathbf{K}_n) defined by

$$\mathbf{K}_n = \sum_{\ell=1}^n |\mathbb{T}_\ell^*|^{-1} (\Delta \mathbf{M}_{\ell+1} \Delta \mathbf{M}_{\ell+1}^t - \mathbb{E}[\Delta \mathbf{M}_{\ell+1} \Delta \mathbf{M}_{\ell+1}^t \mid \mathcal{F}_\ell^\mathcal{O}])$$

is square integrable and its component-wise increasing processes are $\mathcal{O}(n)$ a.s. on $\bar{\mathcal{E}}$.

Let (Ξ_n) be a sequence of $p \times p$ symmetric positive definite matrices, adapted to $\mathbb{F}^\mathcal{O}$, such that

(A.4) On $\bar{\mathcal{E}}$, $|\mathbb{T}_n^*|^{-1} \Xi_n$ converges a.s. to a positive definite matrix Ξ .

Then $\mathbf{M}_n^t \Xi_{n-1}^{-1} \mathbf{M}_n = \mathcal{O}(n)$ and $\|\mathbf{M}_n\|^2 = \mathcal{O}(nm^n)$ a.s. on $\bar{\mathcal{E}}$. If in addition, the entries of (\mathbf{M}_n) satisfy

(A.5) $\sup_n \mathbb{E}[(m^{-n/2} \sum_{k \in \mathbb{G}_n^*} w_k^i)^4 \mid \mathcal{F}_{n-1}^\mathcal{O}] < \infty$ a.s.,

then for all $\delta > 1/2$, $\|\mathbf{M}_n\|^2 = o(n^\delta m^n)$ a.s. and

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \Xi_{\ell-1}^{-1} \mathbf{M}_\ell = \text{tr}(\Gamma \Xi^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \quad (5.1)$$

Proof of the first part of Theorem 5.1 From now on, let us suppose that we are on the non-extinction set $\bar{\mathcal{E}}$. As Ξ_n is definite positive, it is invertible. For all $n \geq 1$, denote $\mathcal{V}_n = \mathbf{M}_n^t \Xi_{n-1}^{-1} \mathbf{M}_n$. We have

$$\begin{aligned} \mathcal{V}_{n+1} &= \mathbf{M}_{n+1}^t \Xi_n^{-1} \mathbf{M}_{n+1} = (\mathbf{M}_n + \Delta \mathbf{M}_{n+1})^t \Xi_n^{-1} (\mathbf{M}_n + \Delta \mathbf{M}_{n+1}), \\ &= \mathcal{V}_n - \mathbf{M}_n^t (\Xi_{n-1}^{-1} - \Xi_n^{-1}) \mathbf{M}_n + 2 \mathbf{M}_n^t \Xi_n^{-1} \Delta \mathbf{M}_{n+1} + \Delta \mathbf{M}_{n+1}^t \Xi_n^{-1} \Delta \mathbf{M}_{n+1}, \end{aligned}$$

since $\mathbf{M}_n^t \Xi_n^{-1} \Delta \mathbf{M}_{n+1}$ is scalar, and hence equal to its own transpose. By summing over the identity above, we obtain

$$\mathcal{V}_{n+1} + \mathcal{A}_n = \mathcal{V}_1 + \mathcal{B}_{n+1} + \mathcal{W}_{n+1}, \quad (5.2)$$

$$\text{where} \quad \mathcal{A}_n = \sum_{\ell=1}^n \mathbf{M}_\ell^t (\Xi_{\ell-1}^{-1} - \Xi_\ell^{-1}) \mathbf{M}_\ell, \quad (5.3)$$

$$\mathcal{B}_{n+1} = 2 \sum_{\ell=1}^n \mathbf{M}_\ell^t \Xi_\ell^{-1} \Delta \mathbf{M}_{\ell+1} \quad \text{and} \quad \mathcal{W}_{n+1} = \sum_{\ell=1}^n \Delta \mathbf{M}_{\ell+1}^t \Xi_\ell^{-1} \Delta \mathbf{M}_{\ell+1}.$$

The asymptotic behavior of $\mathcal{V}_n = \mathbf{M}_n^t \Xi_{n-1}^{-1} \mathbf{M}_n$ is obtained in three steps through the study of \mathcal{W}_n , \mathcal{B}_n and \mathcal{A}_n respectively.

Step 1: Asymptotic behavior of \mathcal{W}_n . Let us prove the following convergence

$$\frac{1}{n} \mathcal{W}_n = \frac{m-1}{m} \text{tr}(\Gamma \Xi^{-1}) \quad \text{a.s. on } \bar{\mathcal{E}} \quad (5.4)$$

Recall that we work on $\bar{\mathcal{E}}$. Rewrite \mathcal{W}_{n+1} as

$$\mathcal{W}_{n+1} = \sum_{\ell=1}^n (|\mathbb{T}_\ell^*|^{-1/2} \Delta \mathbf{M}_{\ell+1})^t (|\mathbb{T}_\ell^*| \Xi_\ell^{-1}) (|\mathbb{T}_\ell^*|^{-1/2} \Delta \mathbf{M}_{\ell+1}).$$

We want to apply the following Lemma (which proof is postponed) for $\Delta_n = \Xi_n^{-1} |\mathbb{T}_n^*|$ and $\mathbf{Z}_n = |\mathbb{T}_n^*|^{-1/2} \Delta \mathbf{M}_{n+1}$.

Lemma 5.2 *Let (Δ_n) be a sequence of $d \times d$ real positive semidefinite symmetric matrices and (Z_n) a sequence of \mathbb{R}^d , such that $\lim_{n \rightarrow \infty} \Delta_n = \Delta$ with Δ definite positive. Then one has*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n Z_\ell^t \Delta_\ell Z_\ell = Z \iff \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n Z_\ell^t \Delta Z_\ell = Z.$$

Indeed, one has $\lim_{n \rightarrow \infty} |\mathbb{T}_\ell^*| \Xi_\ell^{-1} = \Xi^{-1}$ a.s. thanks to (A.4), and Ξ^{-1} is positive definite. Thus, there only remains to prove the a.s. convergence of the real sequence $(n^{-1} \mathcal{T}_{n+1})$, where

$$\begin{aligned} \mathcal{T}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1}^t \Xi^{-1} \Delta M_{\ell+1}}{|\mathbb{T}_\ell^*|} = \text{tr}(\Xi^{-1/2} \mathbf{H}_{n+1} \Xi^{-1/2}) = \text{tr}(\mathbf{H}_{n+1} \Xi^{-1}), \\ \mathbf{H}_{n+1} &= \sum_{\ell=1}^n \frac{\Delta M_{\ell+1} \Delta M_{\ell+1}^t}{|\mathbb{T}_\ell^*|} = \sum_{\ell=1}^n \frac{\Gamma_\ell}{|\mathbb{T}_\ell^*|} + \mathbf{K}_n. \end{aligned}$$

On the one hand, by Assumption (A.3), we know that $\mathbf{K}_n = o(n)$ a.s. on $\bar{\mathcal{E}}$. On the other hand, by definition, one has

$$\frac{\Gamma_\ell}{|\mathbb{T}_\ell^*|} = \left(\frac{\langle \mathbf{M} \rangle_{\ell+1}}{|\mathbb{T}_\ell^*|} - \frac{|\mathbb{T}_{\ell-1}^*| \langle \mathbf{M} \rangle_\ell}{|\mathbb{T}_\ell^*| |\mathbb{T}_{\ell-1}^*|} \right) \xrightarrow{\ell \rightarrow \infty} \left(\Gamma - \frac{\Gamma}{m} \right) = \Gamma \frac{m-1}{m}$$

by Assumption (A.3). Hence, Cesaro convergence yields

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbf{H}_n = \frac{m-1}{m} \Gamma \quad \text{a.s.}$$

We have thus proved the a.s. convergence of $n^{-1} \mathcal{T}_{n+1}$ to $\text{tr}(\Gamma \Xi^{-1})(m-1)/m$. All the assumptions of Lemma 5.2 are satisfied, which leads us to (5.4).

Step 2: Let us prove that $\langle \mathcal{B} \rangle_{n+1} = \mathcal{O}(1 + \mathcal{A}_n)$ a.s. on $\bar{\mathcal{E}}$.

By definition, the process (\mathcal{B}_n) is a scalar $\mathbb{F}^{\mathcal{O}}$ -martingale, and one clearly has

$$\mathbb{E}[\Delta \mathcal{B}_{n+1}^2 | \mathcal{F}_n^{\mathcal{O}}] = 4 \mathbf{M}_n^t \Xi_n^{-1} \Gamma_n \Xi_n^{-1} \mathbf{M}_n \quad \text{a.s.} \quad (5.5)$$

Let us prove that there exists some $\alpha > 0$, some $n_1 \geq 1$, such that for $n \geq n_1$,

$$\Xi_n^{-1} \Gamma_n \Xi_n^{-1} \leq \alpha (\Xi_{n-1}^{-1} - \Xi_n^{-1}), \quad (5.6)$$

in the sense that the symmetric matrix $\alpha (\Xi_{n-1}^{-1} - \Xi_n^{-1}) - \Xi_n^{-1} \Gamma_n \Xi_n^{-1}$ is positive semidefinite. If we remark that this matrix equals

$$|\mathbb{T}_n^*|^{-1} \left(\alpha (|\mathbb{T}_n^*| \Xi_{n-1}^{-1} - |\mathbb{T}_n^*| \Xi_n^{-1}) - |\mathbb{T}_n^*| \Xi_n^{-1} |\mathbb{T}_n^*|^{-1} \Gamma_n |\mathbb{T}_n^*| \Xi_n^{-1} \right),$$

it is clear that it converges, as n goes to ∞ , to $(\alpha m \Xi^{-1} - \Xi^{-1} \Gamma \Xi^{-1})(m-1)/m$. For $u \in \mathbb{R}^p$, one has

$$\begin{aligned} u^t (\alpha m \Xi^{-1} - \Xi^{-1} \Gamma \Xi^{-1}) u &= \alpha m u^t \Xi^{-1} u - u^t \Xi^{-1} \Gamma \Xi^{-1} u \\ &\geq \left(\alpha m \lambda_-(\Xi^{-1}) - \lambda_+(\Xi^{-1} \Gamma \Xi^{-1}) \right) u^t u, \end{aligned}$$

where $\lambda_+(\mathbf{\Delta})$ and $\lambda_-(\mathbf{\Delta})$ denote respectively the greatest and smallest eigenvalues of $\mathbf{\Delta}$. Since $\mathbf{\Xi}^{-1}$ is positive definite by Assumption **(A.4)**, $\lambda_-(\mathbf{\Xi}^{-1}) > 0$, so that we can find $\alpha > 0$ such that $\alpha m \lambda_-(\mathbf{\Xi}^{-1}) - \lambda_+(\mathbf{\Xi}^{-1} \mathbf{\Gamma} \mathbf{\Xi}^{-1}) > 0$. This last inequality yields that $(\alpha m \mathbf{\Xi}^{-1} - \mathbf{\Xi}^{-1} \mathbf{\Gamma} \mathbf{\Xi}^{-1})(m-1)/m$ is a symmetric positive definite matrix, hence so are $\alpha(\mathbf{\Xi}_{n-1}^{-1} - \mathbf{\Xi}_n^{-1}) - \mathbf{\Xi}_n^{-1} \mathbf{\Gamma}_n \mathbf{\Xi}_n^{-1}$ for large enough n . Plugging the majoration given by Eq. (5.6) into Eq. (5.5) then yields

$$\begin{aligned} \langle \mathcal{B} \rangle_{n+1} &\leq 4 \left(\sum_{k=1}^{n_1-1} \mathbf{M}_k^t (\mathbf{\Xi}_k^{-1} \mathbf{\Gamma}_k \mathbf{\Xi}_k^{-1} - \alpha (\mathbf{\Xi}_{k-1}^{-1} - \mathbf{\Xi}_k^{-1})) \mathbf{M}_k \right) + 4\alpha \mathcal{A}_n \\ &\leq c(1 + \mathcal{A}_n), \end{aligned}$$

with c some constant with respect to n . This ends Step 2.

Step 3: Asymptotic behavior of \mathcal{V}_n and \mathbf{M}_n .

Thanks to Eq. (5.6) and the fact that $\mathbf{\Xi}_n^{-1} \mathbf{\Gamma}_n \mathbf{\Xi}_n^{-1}$ is positive semidefinite, (\mathcal{A}_n) is nonnegative and increasing for large enough n , so that either it is convergent, or it goes to infinity. In the former case, Step 2 with the law of large numbers for real martingales yield that (\mathcal{B}_n) also converges, and Step 1 immediately gives $\mathcal{V}_{n+1} = \mathcal{O}(n)$ a.s. on $\bar{\mathcal{E}}$ by Eq. (5.2). In the latter case, Step 2 again with the law of large numbers for real martingales yield that $\mathcal{B}_n = o(\mathcal{A}_n)$. We then deduce from decomposition (5.2) and Step 1 that $\mathcal{V}_{n+1} + \mathcal{A}_n = o(\mathcal{A}_n) + \mathcal{O}(n)$ a.s. on $\bar{\mathcal{E}}$ leading to $\mathcal{V}_{n+1} = \mathcal{O}(n)$ a.s. on $\bar{\mathcal{E}}$, since both \mathcal{A}_n and \mathcal{V}_n are non-negative.

We are now able to prove the first part of Theorem 5.1. By definition of \mathcal{V}_n , we have directly $\mathcal{V}_n = \mathbf{M}_n^t \mathbf{\Xi}_{n-1}^{-1} \mathbf{M}_n = \mathcal{O}(n)$. Moreover, the matrices $\mathbf{\Xi}_n$ being positive definite, we have

$$\|\mathbf{M}_n\|^2 = \mathbf{M}_n^t \mathbf{M}_n \leq \mathbf{M}_n^t \mathbf{\Xi}_{n-1}^{-1} \mathbf{M}_n (\lambda_-(\mathbf{\Xi}_{n-1}^{-1}))^{-1}$$

on $\bar{\mathcal{E}}$. Finally, Assumption **(A.4)** and convergence (2.4) yield $\|\mathbf{M}_n\|^2 = \mathcal{O}(nm^n)$ a.s., which completes the proof of the first part of Theorem 5.1. \square

Proof of Lemma 5.2 Recall that if $\mathbf{\Delta}$ is a symmetric matrix, then for any $u \in \mathbb{R}^d$, $\lambda_-(\mathbf{\Delta}) u^t u \leq u^t \mathbf{\Delta} u \leq \lambda_+(\mathbf{\Delta}) u^t u$. Applying this result for $\mathbf{\Delta}_\ell - \mathbf{\Delta}$ and $\mathbf{\Delta}$ we obtain

$$|\mathbf{Z}_\ell^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{Z}_\ell| \leq \max\{|\lambda_+(\mathbf{\Delta}_\ell - \mathbf{\Delta})|, |\lambda_-(\mathbf{\Delta}_\ell - \mathbf{\Delta})|\} \mathbf{Z}_\ell^t \mathbf{Z}_\ell \quad (5.7)$$

$$\leq \frac{\max\{|\lambda_+(\mathbf{\Delta}_\ell - \mathbf{\Delta})|, |\lambda_-(\mathbf{\Delta}_\ell - \mathbf{\Delta})|\}}{\lambda_-(\mathbf{\Delta})} \mathbf{Z}_\ell^t \mathbf{\Delta} \mathbf{Z}_\ell. \quad (5.8)$$

By assumption, $\max\{|\lambda_+(\mathbf{\Delta}_\ell - \mathbf{\Delta})|, |\lambda_-(\mathbf{\Delta}_\ell - \mathbf{\Delta})|\}$ converges to 0 as ℓ goes to ∞ , so that for any $\varepsilon > 0$, one has

$$\frac{1}{n} \sum_{\ell=1}^n |\mathbf{Z}_\ell^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{Z}_\ell| \leq \frac{c_\varepsilon}{n} + \varepsilon \frac{1}{n} \sum_{\ell=1}^n \mathbf{Z}_\ell^t \mathbf{\Delta} \mathbf{Z}_\ell,$$

for some c_ε constant with respect to n . If $\frac{1}{n} \sum_{\ell=1}^n \mathbf{Z}_\ell^t \Delta \mathbf{Z}_\ell$ converges to Z , letting n go to ∞ leads to

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n |\mathbf{Z}_\ell^t (\Delta_\ell - \Delta) \mathbf{Z}_\ell| \leq \varepsilon Z,$$

and this holding for any $\varepsilon > 0$, we obtain that $\frac{1}{n} \sum_{\ell=1}^n \mathbf{Z}_\ell^t (\Delta_\ell - \Delta) \mathbf{Z}_\ell$ converges to 0, and therefore $\frac{1}{n} \sum_{\ell=1}^n \mathbf{Z}_\ell^t \Delta_\ell \mathbf{Z}_\ell$ converges to Z . The same arguments, replacing Δ by Δ_ℓ in Eq. (5.8) yields the reverse implication. \square

Proof of the second part of Theorem 5.1 Let us rewrite each entry M_n^q of the martingale \mathbf{M}_n as

$$M_n^q = \sum_{\ell=1}^n m^{\ell/2} \frac{1}{m^{\ell/2}} \sum_{k \in \mathbb{G}_\ell^*} w_k^q = \sum_{\ell=1}^n m^{\ell/2} x_\ell^q.$$

We apply Wei's lemma [27, p 1672] to the martingale difference sequence (x_ℓ^q) and the scalar sequence $(m^{\ell/2})$ for the function $f(x) = (\log x)^{\delta/2}$ with $\delta > 1/2$ using Assumption **(A.5)**. One obtains $M_n^q = o(m^{n/2} n^{\delta/2})$. As M_n^q is the q -th entry of \mathbf{M}_n , one has $\|\mathbf{M}_n\|^2 = o(n^\delta m^n)$ a.s. Now recall that $\mathcal{V}_n = \mathbf{M}_n^t \Xi_{n-1}^{-1} \mathbf{M}_n$, therefore, one has

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathcal{V}_n = \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \mathbf{M}_n^t \Xi_{n-1}^{-1} \mathbf{M}_n = o(n^\delta) \quad \text{a.s.}$$

for all $\delta > 1/2$. In particular, for $\delta = 1$, $\mathcal{V}_n = o(n)$. Then, Eq. (5.2) together with the proof of the first part of Theorem 5.1, and the law of large numbers for real martingales, yield

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\mathcal{A}_n}{n} = \frac{m-1}{m} \text{tr}(\Gamma \Xi^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.} \quad (5.9)$$

Henceforth, we work on $\bar{\mathcal{E}}$. Rewrite \mathcal{A}_n as

$$\mathcal{A}_n = \sum_{\ell=1}^n \mathbf{M}_\ell^t (\Xi_{\ell-1}^{-1} - \Xi_\ell^{-1}) \mathbf{M}_\ell = \sum_{\ell=1}^n (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t \Delta_\ell (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell),$$

where $\Delta_\ell = \mathbf{I}_4 - \Xi_{\ell-1}^{1/2} \Xi_\ell^{-1} \Xi_{\ell-1}^{1/2}$. To end the proof, we use Lemma 5.2 with $\mathbf{Z}_\ell = \Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell$. Thanks to Assumption **(A.4)** we have $\lim_{\ell \rightarrow \infty} \Delta_\ell = \frac{m-1}{m} \mathbf{I}_4$, so that Δ_ℓ and its limit are definite positive for large enough ℓ as $m > 1$. Besides, Eq. (5.9) gives the convergence of $\frac{1}{n} \sum_{\ell=1}^n (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t \Delta_\ell (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell)$. Applying Lemma 5.2 thus gives

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t \frac{m-1}{m} \mathbf{I}_4 (\Xi_{\ell-1}^{-1/2} \mathbf{M}_\ell) = \frac{m-1}{m} \text{tr}(\Gamma \Xi^{-1}),$$

which is the expected result. \square

We now state a corollary to Theorem 5.1 that will be useful in the sequel when dealing with indefinite symmetric matrices.

Corollary 5.3 *Let (\mathbf{M}_n) and $(\mathbf{\Xi}_n)$ satisfying the assumptions of Theorem 5.1. Let $(\mathbf{\Delta}_n)$ be a sequence of invertible symmetric matrices converging almost surely on $\bar{\mathcal{E}}$ to an invertible symmetric matrix $\mathbf{\Delta}$. Then one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{\Delta}_\ell \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell = \text{tr}(\mathbf{\Gamma} \mathbf{\Xi}^{-1} \mathbf{\Delta}) \mathbb{1}_{\bar{\mathcal{E}}}.$$

Proof The first step is analogous to Lemma 5.2: we want to replace $\mathbf{\Delta}_\ell$ by the fixed matrix $\mathbf{\Delta}$ in the equation above. As $\mathbf{\Delta}$ is possibly indefinite, we cannot directly apply Lemma 5.2. We modify its proof as follow, starting from Eq. (5.7)

$$\begin{aligned} & |(\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell| \\ & \leq \max\{|\lambda_+(\mathbf{\Delta}_\ell - \mathbf{\Delta})|, |\lambda_-(\mathbf{\Delta}_\ell - \mathbf{\Delta})|\} \mathbf{M}_\ell^t \mathbf{\Xi}_{\ell-1}^{-1} \mathbf{M}_\ell. \end{aligned}$$

By assumption, $\max\{|\lambda_+(\mathbf{\Delta}_\ell - \mathbf{\Delta})|, |\lambda_-(\mathbf{\Delta}_\ell - \mathbf{\Delta})|\}$ converges to 0 as ℓ goes to ∞ , so that for any $\varepsilon > 0$, one has

$$\frac{1}{n} \sum_{\ell=1}^n |(\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell| \leq \frac{c_\varepsilon}{n} + \varepsilon \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \mathbf{\Xi}_{\ell-1}^{-1} \mathbf{M}_\ell,$$

for some c_ε constant with respect to n . As (\mathbf{M}_n) and $(\mathbf{\Xi}_n)$ satisfy the assumptions of Theorem 5.1, Eq.(5.1) yields

$$\limsup_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n |(\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell| \leq \varepsilon \text{tr}(\mathbf{\Gamma} \mathbf{\Xi}^{-1}),$$

and this holding for any $\varepsilon > 0$, we obtain that $\frac{1}{n} \sum_{\ell=1}^n (\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t (\mathbf{\Delta}_\ell - \mathbf{\Delta}) \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell$ converges to 0, and therefore $\frac{1}{n} \sum_{\ell=1}^n (\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t \mathbf{\Delta}_\ell \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell$ has the same limit as $\frac{1}{n} \sum_{\ell=1}^n (\mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell)^t \mathbf{\Delta} \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell$. Thus, it only remains to prove the convergence of $\frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{\Delta} \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell$. Once again Theorem 5.1 cannot be directly applied with the sequence $(\mathbf{\Xi}_{\ell-1}^{1/2} \mathbf{\Delta}^{-1} \mathbf{\Xi}_{\ell-1}^{1/2})$ as these matrices may be indefinite. To get around this difficulty, recall that any symmetric matrix can be rewrite as a difference between two symmetric definite positive matrices. Thus, set $\mathbf{\Delta} = \mathbf{\Delta}_+ - \mathbf{\Delta}_-$, where $\mathbf{\Delta}_+$ and $\mathbf{\Delta}_-$ are symmetric positive definite. Applying twice Theorem 5.1 with the sequences $(\mathbf{\Xi}_{\ell-1}^{1/2} \mathbf{\Delta}_+^{-1} \mathbf{\Xi}_{\ell-1}^{1/2})$ and $(\mathbf{\Xi}_{\ell-1}^{1/2} \mathbf{\Delta}_-^{-1} \mathbf{\Xi}_{\ell-1}^{1/2})$, and using the linearity of the trace, yields a.s. on $\bar{\mathcal{E}}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{\Delta} \mathbf{\Xi}_{\ell-1}^{-1/2} \mathbf{M}_\ell = \text{tr}(\mathbf{\Gamma} \mathbf{\Xi}^{-1/2} (\mathbf{\Delta}_+ - \mathbf{\Delta}_-) \mathbf{\Xi}^{-1/2}) = \text{tr}(\mathbf{\Gamma} \mathbf{\Xi}^{-1} \mathbf{\Delta}),$$

ending the proof. \square

5.2 Rate of convergence for $\hat{\boldsymbol{\theta}}_n$

We apply Theorem 5.1 to a suitably chosen martingale. We have

$$\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \mathbf{S}_{n-1}^{-1} \sum_{k \in \mathbb{T}_{n-1}} (\epsilon_{2k}, X_k \epsilon_{2k}, \epsilon_{2k+1}, X_k \epsilon_{2k+1})^t = \mathbf{S}_{n-1}^{-1} \mathbf{M}_n, \quad (5.10)$$

where

$$\mathbf{M}_n = \sum_{k \in \mathbb{T}_{n-1}} (\epsilon_{2k}, X_k \epsilon_{2k}, \epsilon_{2k+1}, X_k \epsilon_{2k+1})^t. \quad (5.11)$$

Under assumptions **(H.1-3)**, we have for all $n \geq 0$, $k \in \mathbb{G}_n$, $\mathbb{E}[\epsilon_{2k+i} | \mathcal{F}_n^{\mathcal{O}}] = \mathbb{E}[X_k \epsilon_{2k+i} | \mathcal{F}_n^{\mathcal{O}}] = 0$ and (\mathbf{M}_n) is a square integrable $(\mathcal{F}_n^{\mathcal{O}})$ -martingale, so that Assumption **(A.1)** holds. Set $\mathbf{\Gamma}_\ell = \mathbb{E}[\Delta \mathbf{M}_{\ell+1} \Delta \mathbf{M}_{\ell+1}^t | \mathcal{F}_\ell^{\mathcal{O}}]$. Thus one has

$$\langle \mathbf{M} \rangle_n = \sum_{\ell=0}^{n-1} \mathbf{\Gamma}_\ell = \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_\ell} \gamma_k \otimes \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix} = \sum_{k \in \mathbb{T}_{n-1}} \gamma_k \otimes \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix},$$

where

$$\gamma_k = \begin{pmatrix} \delta_{2k}(\sigma_\varepsilon^2 + 2X_k \rho_{00} + X_k^2 \sigma_\eta^2) & \delta_{2k} \delta_{2k+1}(\rho_\varepsilon + 2X_k \rho + X_k^2 \rho_\eta) \\ \delta_{2k} \delta_{2k+1}(\rho_\varepsilon + 2X_k \rho + X_k^2 \rho_\eta) & \delta_{2k+1}(\sigma_\varepsilon^2 + 2X_k \rho_{11} + X_k^2 \sigma_\eta^2) \end{pmatrix}. \quad (5.12)$$

Lemma 5.4 *Under assumptions **(H.1-5)** and if $\kappa \geq 4$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\langle \mathbf{M} \rangle_n}{|\mathbb{T}_{n-1}^*|} = \mathbf{\Gamma} \mathbb{1}_{\bar{\mathcal{E}}} = \begin{pmatrix} \mathbf{\Gamma}^0 & \mathbf{\Gamma}^{01} \\ \mathbf{\Gamma}^{01} & \mathbf{\Gamma}^1 \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.,$$

where $\mathbf{\Gamma}^0$, $\mathbf{\Gamma}^{01}$ and $\mathbf{\Gamma}^1$ are the 2×2 matrices defined by

$$\begin{aligned} \mathbf{\Gamma}^i &= \begin{pmatrix} \sigma_\varepsilon^2 \ell_i(0) + 2\rho_{ii} \ell_i(1) + \sigma_\eta^2 \ell_i(2) & \sigma_\varepsilon^2 \ell_i(1) + 2\rho_{ii} \ell_i(2) + \sigma_\eta^2 \ell_i(3) \\ \sigma_\varepsilon^2 \ell_i(1) + 2\rho_{ii} \ell_i(2) + \sigma_\eta^2 \ell_i(3) & \sigma_\varepsilon^2 \ell_i(2) + 2\rho_{ii} \ell_i(3) + \sigma_\eta^2 \ell_i(4) \end{pmatrix}, \\ \mathbf{\Gamma}^{01} &= \begin{pmatrix} \rho_\varepsilon \ell_{01}(0) + 2\rho \ell_{01}(1) + \rho_\eta \ell_{01}(2) & \rho_\varepsilon \ell_{01}(1) + 2\rho \ell_{01}(2) + \rho_\eta \ell_{01}(3) \\ \rho_\varepsilon \ell_{01}(1) + 2\rho \ell_{01}(2) + \rho_\eta \ell_{01}(3) & \rho_\varepsilon \ell_{01}(2) + 2\rho \ell_{01}(3) + \rho_\eta \ell_{01}(4) \end{pmatrix}. \end{aligned}$$

Proof This is a direct consequence of Proposition 4.11. \square

Hence, Assumption **(A.2)** holds if $\kappa \geq 4$. The process (\mathbf{K}_n) is clearly a square integrable martingale if $\gamma \geq 2$. It is not difficult to check that its component-wise increasing process is at most of the order of

$$\sum_{\ell=1}^n \frac{1}{|\mathbb{T}_\ell^*|^2} \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} X_k^8, \quad (5.13)$$

with $i \in \{0, 1\}$. Proposition 4.11 ensures that $|\mathbb{T}_n^*|^{-1} \sum_{k \in \mathbb{G}_n} \delta_{2k+i} X_k^4$ converges a.s. on $\bar{\mathcal{E}}$ provided $\kappa \geq 4$, it is therefore bounded by some constant C . As a result, its square is also bounded by C^2 and $|\mathbb{T}_\ell^*|^{-2} \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} X_k^8 \leq C^2$ a.s. on $\bar{\mathcal{E}}$. Finally, we get that

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \sum_{\ell=1}^n \frac{1}{|\mathbb{T}_\ell^*|^2} \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} X_k^8 \leq C^2 n \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}},$$

so that Assumption **(A.3)** holds if $\kappa \geq 4$. We now introduce a new sequence of matrices $\mathbf{\Sigma}_n$. They are defined as a standardized version of the increasing process

of (\mathbf{M}_n) , with the variance coefficients σ_ε^2 and σ_η^2 set to 1 and all the covariance coefficients $\rho_\varepsilon, \rho_\eta, \rho_{ij}$ set to 0, namely

$$\Sigma_n = \sum_{k \in \mathbb{T}_n} (1 + X_k^2) \begin{pmatrix} \delta_{2k} & 0 \\ 0 & \delta_{2k+1} \end{pmatrix} \otimes \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix},$$

Note that Σ_n is definite positive as soon as the X_k^2 are not constant. The next result is again a direct consequence of Proposition 4.11.

Lemma 5.5 *Under assumptions (H.1-5) and if $\kappa \geq 4$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{\Sigma_n}{|\mathbb{T}_n^*|} = \Sigma \mathbb{1}_{\bar{\mathcal{E}}} = \begin{pmatrix} \Sigma^0 & 0 \\ 0 & \Sigma^1 \end{pmatrix} \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.,$$

where Σ^i is the 2×2 matrix $\Sigma^i = \begin{pmatrix} \ell_i(0) + \ell_i(2) & \ell_i(1) + \ell_i(3) \\ \ell_i(1) + \ell_i(3) & \ell_i(2) + \ell_i(4) \end{pmatrix}$. In addition, Σ is definite positive.

Thus, Assumption (A.4) also holds if $\kappa \geq 4$.

Lemma 5.6 *Under assumptions (H.1-5) and if $\kappa \geq 4$, for $i \in \{0, 1\}$ and $q \in \{0, 1\}$, one has*

$$\sup_n \left\{ m^{-2n} \mathbb{E} \left[\left(\sum_{k \in \mathbb{G}_n^*} X_k^q \epsilon_{2k+i} \right)^4 \mid \mathcal{F}_n^{\mathcal{O}} \right] \right\} < \infty \quad a.s.$$

Proof The left hand of the equation above is at most of the order of

$$\sup_n \left\{ \left(\frac{1}{m^n} \sum_{k \in \mathbb{G}_n^*} \delta_{2k+i} X_k^4 \right)^2 + \frac{1}{m^{2n}} \sum_{k \in \mathbb{G}_n^*} \delta_{2k+i} X_k^8 \right\}.$$

We bound these terms along the sakes lines as Eq. (5.13). □

We have now proved that Assumptions (A.1-5) of Theorem 5.1 hold for the martingale (\mathbf{M}_n) and the sequence of definite positive matrices $(\Xi_n = \Sigma_n)$, thus we obtain the following result.

Proposition 5.7 *Under assumptions (H.1-5) and if $\kappa \geq 4$, one has*

$$\mathbf{M}_n^t \Sigma_{n-1}^{-1} \mathbf{M}_n = \mathcal{O}(n), \quad \text{and} \quad \|\mathbf{M}_n\|^2 = \mathcal{O}(nm^n) \quad a.s.$$

In addition, for all $\delta > 1/2$, $\|\mathbf{M}_n\|^2 = o(n^\delta m^n)$ a.s. and

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{\ell=1}^n \mathbf{M}_\ell^t \Sigma_{\ell-1}^{-1} \mathbf{M}_\ell = \text{tr}(\Gamma \Sigma^{-1}) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Proof of Theorem 3.2 Recall that $\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta} = \mathbf{S}_{n-1}^{-1} \mathbf{M}_n$, one has

$$\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = \mathbf{M}_n^t \mathbf{S}_{n-1}^{-2} \mathbf{M}_n \leq \|\mathbf{M}_n\|^2 \lambda_+(\mathbf{S}_{n-1}^{-2}),$$

and use Proposition 4.14 to conclude that $\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}\|^2 = o(n^\delta m^{-n})$ a.s. For the quadratic strong law, Proposition 5.7 yields the following a.s. limit on $\bar{\mathcal{E}}$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{\ell=1}^n |\mathbb{T}_{\ell-1}^*| (\hat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta})^t (|\mathbb{T}_{\ell-1}^*|^{-1} \mathbf{S}_{\ell-1} \Sigma_{\ell-1}^{-1} \mathbf{S}_{\ell-1}) (\hat{\boldsymbol{\theta}}_\ell - \boldsymbol{\theta}) = \text{tr}(\Gamma \Sigma^{-1})$$

and the result is obtained using Proposition 4.14, Lemmas 5.5 and 5.2. □

5.3 Rate of convergence for $\widehat{\boldsymbol{\sigma}}_n$

We now turn to the convergence of $\widehat{\boldsymbol{\sigma}}_n - \boldsymbol{\sigma}_n$. Recall that $\widehat{\boldsymbol{\sigma}}_n$ defined in Eq. (3.1) and $\boldsymbol{\sigma}_n$ in Eq. (3.3). Note that

$$\mathbf{U}_{n-1}(\widehat{\boldsymbol{\sigma}}_n - \boldsymbol{\sigma}_n) = \mathbf{P}_n^\sigma + 2\mathbf{R}_n^\sigma, \quad (5.14)$$

with

$$\mathbf{P}_n^\sigma = \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} (\widehat{\epsilon}_{2k} - \epsilon_{2k})^2 + (\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1})^2 \\ 2X_k(\widehat{\epsilon}_{2k} - \epsilon_{2k})^2 \\ 2X_k(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1})^2 \\ X_k^2((\widehat{\epsilon}_{2k} - \epsilon_{2k})^2 + (\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1})^2) \end{pmatrix},$$

$$\mathbf{R}_n^\sigma = \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \epsilon_{2k}(\widehat{\epsilon}_{2k} - \epsilon_{2k}) + \epsilon_{2k+1}(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) \\ 2X_k\epsilon_{2k}(\widehat{\epsilon}_{2k} - \epsilon_{2k}) \\ 2X_k\epsilon_{2k+1}(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) \\ X_k^2(\epsilon_{2k}(\widehat{\epsilon}_{2k} - \epsilon_{2k}) + \epsilon_{2k+1}(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1})) \end{pmatrix}.$$

We are going to study separately the asymptotic properties of \mathbf{P}_n^σ and \mathbf{R}_n^σ .

Lemma 5.8 *Under assumptions (H.1-5) and if $\kappa \geq 4$, for all $i \in \{0, 1\}$ and $p \in \{0, 1, 2\}$, we have the following convergences*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_n} X_k^p (\widehat{\epsilon}_{2k+i} - \epsilon_{2k+i})^2 &= q_i(p) \mathbb{1}_{\overline{\mathcal{E}}} \\ &= (m-1) \text{tr}(\boldsymbol{\Gamma}^i (\mathbf{S}^i)^{-2} \mathbf{S}^i(p)) \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.} \end{aligned}$$

where $\mathbf{S}^i(p)$ is the 2×2 matrix defined by $\mathbf{S}^i(p) = \begin{pmatrix} \ell_i(p) & \ell_i(p+1) \\ \ell_i(p+1) & \ell_i(p+2) \end{pmatrix}$.

Proof We detail the proof for $i = 0$, the same arguments holding for $i = 1$, *mutatis mutandis*. We apply Theorem 5.1 to (\mathbf{M}_n^0) , the 2-component vector corresponding to the first two entries of the martingale (\mathbf{M}_n) defined by Eq. (5.11). Let $\boldsymbol{\theta}^0 = (a, b)^t$, $\widehat{\boldsymbol{\theta}}_n^0 = (\widehat{a}_n, \widehat{b}_n)^t$. Clearly, one has $(\widehat{\boldsymbol{\theta}}_n^0 - \boldsymbol{\theta}^0) = (\mathbf{S}_{n-1}^0)^{-1} \mathbf{M}_n^0$, and thus, if $\mathbf{S}_\ell^i(p) = \sum_{k \in \mathbb{T}_\ell} \delta_{2k+i} \begin{pmatrix} X_k^p & X_k^{p+1} \\ X_k^{p+1} & X_k^{p+2} \end{pmatrix}$, one has

$$\begin{aligned} \frac{1}{n} \sum_{k \in \mathbb{T}_n} X_k^p (\widehat{\epsilon}_{2k} - \epsilon_{2k})^2 &= \frac{1}{n} \sum_{\ell=1}^n \sum_{k \in \mathbb{G}_\ell} \delta_{2k} (\widehat{\boldsymbol{\theta}}_\ell^0 - \boldsymbol{\theta}^0)^t \begin{pmatrix} X_k^p & X_k^{p+1} \\ X_k^{p+1} & X_k^{p+2} \end{pmatrix} (\widehat{\boldsymbol{\theta}}_\ell^0 - \boldsymbol{\theta}^0) \\ &= \frac{1}{n} \sum_{\ell=1}^n (\mathbf{M}_\ell^0)^t \boldsymbol{\Delta}_{\ell-1}^{-1}(p) \mathbf{M}_\ell^0, \end{aligned}$$

where $\boldsymbol{\Delta}_\ell(p) = \mathbf{S}_\ell^0 (\mathbf{S}_{\ell+1}^0(p) - \mathbf{S}_\ell^0(p))^{-1} \mathbf{S}_\ell^0$ is a 2×2 symmetric definite positive matrix. Since $p \leq 2 \leq \kappa$, Proposition 4.11 yields a.s. the convergence of $|\mathbb{T}_\ell^*|^{-1} \mathbf{S}_\ell^i(p)$ to $\mathbf{S}^i(p)$ on the non extinction set, and thus of $|\mathbb{T}_\ell|^{-1} \boldsymbol{\Delta}_\ell(p)$ to the symmetric positive definite matrix $\boldsymbol{\Delta}(p) = (m-1)(\mathbf{S}^0)^{-1} \mathbf{S}^0(p) (\mathbf{S}^0)^{-1}$. We can thus apply Theorem 5.1 with the martingale (\mathbf{M}_n^0) and the sequence of symmetric positive definite matrices

$(\Xi_\ell = \Delta_\ell(p))$ satisfying Assumption (A.4). Assumptions (A.1-3) and (A.5) also clearly hold, because (\mathbf{M}_n^0) inherits them from (\mathbf{M}_n) . We thus obtain

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \sum_{k \in \mathbb{T}_n} X_k^p (\widehat{\epsilon}_{2k} - \epsilon_{2k})^2 = (m-1) \text{tr}(\Gamma^0(\mathbf{S}^0)^{-2} \mathbf{S}^0(p)) \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s.$$

which completes the proof. \square

As a consequence, we obtain the a.s. convergence of the sequence (\mathbf{P}_n^σ) .

Lemma 5.9 *Under assumptions (H.1-5) and if $\kappa \geq 4$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathbf{P}_n^\sigma = (q_0(0) + q_1(0), 2q_0(1), 2q_1(1), q_0(2) + q_1(2))^t \mathbb{1}_{\overline{\mathcal{E}}} \quad a.s.$$

It remains to give the limit of the sequence (\mathbf{R}_n^σ) .

Lemma 5.10 *Under assumptions (H.1-5) and if $\kappa \geq 8$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathbf{R}_n^\sigma = 0 \quad a.s.$$

Proof : It is sufficient to prove that (\mathbf{R}_n^σ) is a martingale and that its increasing process is almost surely $\mathcal{O}(n)$. For all $k \in \mathbb{G}_n$, one has

$$\mathbb{E}[\epsilon_{2k} (\widehat{\epsilon}_{2k} - \epsilon_{2k}) \mid \mathcal{F}_n^\mathcal{O}] = \delta_{2k} ((a - \widehat{a}_n) + (b - \widehat{b}_n) X_k) \mathbb{E}[\epsilon_{2k} \mid \mathcal{F}_n^\mathcal{O}] = 0,$$

and we have similar results for the other entries of \mathbf{R}_n^σ . Hence, (\mathbf{R}_n^σ) is a $(\mathcal{F}_n^\mathcal{O})$ -martingale. It is also square integrable. We are going to study (\mathbf{R}_n^σ) component-wise. We give the details for the last entry, the others being treated similarly. For $i \in \{0, 1\}$, set

$$Q_n^i = \sum_{\ell=1}^{n-1} (\boldsymbol{\theta}^i - \widehat{\boldsymbol{\theta}}_\ell^i)^t \sum_{k \in \mathbb{G}_\ell} \delta_{2k+i} \begin{pmatrix} X_k^2 \\ X_k^3 \end{pmatrix} \epsilon_{2k+i}.$$

The last entry of \mathbf{R}_n^σ can then be rewritten as $Q_n^0 + Q_n^1$. The processes (Q_n^i) are clearly $(\mathcal{F}_n^\mathcal{O})$ -martingales with increasing processes equal to

$$\langle Q^i \rangle_n = \sum_{\ell=1}^{n-1} (\mathbf{M}_\ell^i)^t (\mathbf{S}_{\ell-1}^i)^{-1} \boldsymbol{\Delta}_\ell^i (\mathbf{S}_{\ell-1}^i)^{-1} \mathbf{M}_\ell^i,$$

with $\boldsymbol{\Delta}_n^i = \sum_{k \in \mathbb{G}_n} \delta_{2k+i} (\sigma_\varepsilon^2 + 2\rho_{ii} X_k + \sigma_\eta^2 X_k^2) \begin{pmatrix} X_k^4 & X_k^5 \\ X_k^5 & X_k^6 \end{pmatrix}$. Thanks to Proposition 4.11, and since $\kappa \geq 8$, the sequence of matrices $(|\mathbb{T}_n^*|^{-1} \boldsymbol{\Delta}_n^i)$ converges a.s. on the non-extinction set $\overline{\mathcal{E}}$ to a positive definite matrix and hence $|\mathbb{T}_n^*| (\mathbf{S}_{n-1}^i)^{-1} \boldsymbol{\Delta}_n^i (\mathbf{S}_{n-1}^i)^{-1}$ also converges to a definite positive matrix $\boldsymbol{\Delta}^i$. We now apply Theorem 5.1 to (\mathbf{M}_n^i) and $(\mathbf{S}_{n-1}^i (\boldsymbol{\Delta}_n^i)^{-1} \mathbf{S}_{n-1}^i)$ to obtain that $\langle Q^i \rangle_n = \mathcal{O}(n)$, and thus $Q_n^i = o(n)$. The other entries of (\mathbf{R}_n^σ) are dealt with similarly, yielding the result. \square

Proof of Theorem 3.3 The convergence of $\widehat{\boldsymbol{\sigma}}_n - \boldsymbol{\sigma}_n$ is a direct consequence of Eq. (5.14), Proposition 4.14, Lemmas 5.9 and 5.10. The convergence of $\boldsymbol{\sigma}_n$ to $\boldsymbol{\sigma}$ is then a consequence of this previous convergence and Theorem 3.1. \square

5.4 Rate of convergence for $\widehat{\rho}_n$

We now turn to the convergence of $\widehat{\rho}_n - \rho_n$. We follow the same steps as in Section 5.3. One can rewrite $\mathbf{V}_{n-1}(\widehat{\rho}_n - \rho_n)$ as

$$\mathbf{V}_{n-1}(\widehat{\rho}_n - \rho_n) = \mathbf{P}_n^\rho + \mathbf{R}_n^\rho, \quad (5.15)$$

with

$$\begin{aligned} \mathbf{P}_n^\rho &= \sum_{k \in \mathbb{T}_{n-1}} (\widehat{\epsilon}_{2k} - \epsilon_{2k})(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) (1, 2X_k, X_k^2)^t, \\ \mathbf{R}_n^\rho &= \sum_{k \in \mathbb{T}_{n-1}} (\epsilon_{2k+1}(\widehat{\epsilon}_{2k} - \epsilon_{2k}) + \epsilon_{2k}(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1})) (1, 2X_k, X_k^2)^t. \end{aligned}$$

We are going to study separately the asymptotic properties of \mathbf{P}_n^ρ and \mathbf{R}_n^ρ .

Lemma 5.11 *Under assumptions (H.1-5) and if $\kappa \geq 4$, for all $p \in \{0, 1, 2\}$, the following a.s. convergences hold on the non extinction set $\bar{\mathcal{E}}$*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k \in \mathbb{T}_n} X_k^p (\widehat{\epsilon}_{2k} - \epsilon_{2k})(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) &= q_{01}(p) \\ &= \frac{m-1}{2} \text{tr}(\mathbf{\Gamma} \mathbf{S}^{-2} \mathbf{J}^{01}(p)), \end{aligned}$$

$$\text{with } \mathbf{J}^{01}(p) = \begin{pmatrix} 0 & \mathbf{S}^{01}(p) \\ \mathbf{S}^{01}(p) & 0 \end{pmatrix} \text{ and } \mathbf{S}^{01}(p) = \begin{pmatrix} \ell_{01}(p) & \ell_{01}(p+1) \\ \ell_{01}(p+1) & \ell_{01}(p+2) \end{pmatrix}.$$

Proof : First, notice that for all $k \in \mathbb{G}_n$ and $p \in \{0, 1, 2\}$, one has

$$\begin{aligned} &2X_k^p (\widehat{\epsilon}_{2k} - \epsilon_{2k})(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) \\ &= \delta_{2k} \delta_{2k+1} (\widehat{\theta}_n - \theta)^t \begin{pmatrix} 0 & 0 & X_k^p & X_k^{p+1} \\ 0 & 0 & X_k^{p+1} & X_k^{p+2} \\ X_k^p & X_k^{p+1} & 0 & 0 \\ X_k^{p+1} & X_k^{p+2} & 0 & 0 \end{pmatrix} (\widehat{\theta}_n - \theta). \end{aligned}$$

Hence, using Eq. (5.10), one has

$$2 \sum_{k \in \mathbb{T}_n} X_k^p (\widehat{\epsilon}_{2k} - \epsilon_{2k})(\widehat{\epsilon}_{2k+1} - \epsilon_{2k+1}) = \sum_{\ell=1}^n \mathbf{M}_\ell^t \mathbf{S}_{\ell-1}^{-1} (\mathbf{J}_\ell^{01}(p) - \mathbf{J}_{\ell-1}^{01}(p)) \mathbf{S}_{\ell-1}^{-1} \mathbf{M}_\ell,$$

with

$$\mathbf{J}_n^{01}(p) = \begin{pmatrix} 0 & \mathbf{S}_n^{01}(p) \\ \mathbf{S}_n^{01}(p) & 0 \end{pmatrix} \text{ and } \mathbf{S}_n^{01}(p) = \sum_{k \in \mathbb{T}_n} 2\delta_{2k} \delta_{2k+1} X_k^p \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix}.$$

Set $\Delta_n(p) = \mathbf{S}_n^{-1/2} (\mathbf{J}_{n+1}^{01}(p) - \mathbf{J}_n^{01}(p)) \mathbf{S}_n^{-1/2}$. Thanks to Proposition 4.11, we have the following convergence

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \Delta_n(p) = \Delta(p) \mathbb{1}_{\bar{\mathcal{E}}} = (m-1) \mathbf{S}^{-1/2} \mathbf{J}^{01}(p) \mathbf{S}^{-1/2} \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

We now apply Corollary 5.3 to the martingale (\mathbf{M}_n) and the sequences (\mathbf{S}_n) and (Δ_n) to obtain the result. \square

Thus, one obtains the following limit.

Lemma 5.12 *Under assumptions (H.1-5) and if $\kappa \geq 4$, one has the almost sure convergence*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathbf{P}_n^\sigma = (q_{01}(0), 2q_{01}(1), q_{01}(2))^t \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

The limit of \mathbf{R}_n^ρ is obtained similarly to Lemma 5.10 using Corollary 5.3 again instead of Theorem 5.1.

Lemma 5.13 *Under assumptions (H.1-5) and if $\kappa \geq 8$, one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \frac{1}{n} \mathbf{R}_n^\rho = 0 \quad a.s.$$

Proof of Theorem 3.4 The convergence of $\hat{\rho}_n - \rho_n$ is a direct consequence of Eq. (5.15), Proposition 4.14, Lemmas 5.12 and 5.13. The convergence of ρ_n to ρ is a consequence of this previous convergence and Theorem 3.1. \square

6 Asymptotic normality

To derive the central limit theorems (CLT), we use a CLT for martingales given in [22, Theorem 2.1.9]. To this aim, we use a new filtration: instead of using the generation-wise filtration, we will use the sister pair-wise one. Let

$$\mathcal{G}_p^\mathcal{O} = \mathcal{O} \vee \sigma\{\delta_1 X_1, (\delta_{2k} X_{2k}, \delta_{2k+1} X_{2k+1}), 1 \leq k \leq p\}$$

be the σ -algebra generated by the whole history \mathcal{O} of the GW process and all observed individuals up to the offspring of individual p . Hence $(\epsilon_{2k}, \epsilon_{2k+1})$ is $\mathcal{G}_k^\mathcal{O}$ -measurable. In all the sequel, we will work on the non-extinction probability space $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$ and we denote by $\mathbb{E}_{\bar{\mathcal{E}}}$ the corresponding expectation.

Proof of Theorem 3.5, first step For a fixed integer $n \geq 1$, let us define the $\mathcal{G}_p^\mathcal{O}$ -martingale $(M_p^{(n)})_{\{p \geq 1\}}$ by

$$M_p^{(n)} = \frac{1}{|\mathbb{T}_n^*|^{1/2}} \sum_{k=1}^p D_k \quad \text{with} \quad D_k = (\epsilon_{2k}, X_k \epsilon_{2k}, \epsilon_{2k+1}, X_k \epsilon_{2k+1})^t.$$

Under (H.1-5), D_k is clearly a $\mathcal{G}_k^\mathcal{O}$ -martingale difference sequence. For all n , $\nu_n = |\mathbb{T}_n| = (2^{n+1} - 1)$ is a $(\mathcal{G}_n^\mathcal{O})$ -stopping time. Using Eq. (5.11), we have

$$M_{\nu_n}^{(n)} = \frac{1}{|\mathbb{T}_n^*|^{1/2}} \sum_{k=1}^{|\mathbb{T}_n|} D_k = \frac{1}{|\mathbb{T}_n^*|^{1/2}} M_{n+1}.$$

In order to apply Theorem 2.1.9 of [22] we compute the increasing process of $(M_p^{(n)})$. As the non-extinction set $\bar{\mathcal{E}} \in \mathcal{G}_k^\mathcal{O}$ for all $k \geq 1$, one has

$$\mathbb{E}_{\bar{\mathcal{E}}} [D_k D_k^t | \mathcal{G}_{k-1}^\mathcal{O}] = \mathbb{E} [D_k D_k^t | \mathcal{G}_{k-1}^\mathcal{O}] = \gamma_k \otimes \begin{pmatrix} 1 & X_k \\ X_k & X_k^2 \end{pmatrix},$$

where γ_k is defined in Eq. (5.12). Lemma 5.4 gives the following $\mathbb{P}_{\bar{\mathcal{E}}}$ a.s. limit

$$\frac{1}{|\mathbb{T}_n^*|} \sum_{k \in \mathbb{T}_n^*} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] \xrightarrow{n \rightarrow \infty} \mathbf{\Gamma} \quad \text{a.s.}$$

Therefore, the first assumption of Theorem 2.1.9 of [22] holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. Let us now turn to the second condition. Thanks to Hölder and Tchebycheff inequalities, we have

$$\begin{aligned} & \frac{1}{|\mathbb{T}_n^*|} \sum_{k=1}^{\nu_n} \mathbb{E} \left(\|D_k\|^2 \mathbb{1}_{\{\|M_k^{(n)} - M_{k-1}^{(n)}\| > \epsilon\}} \mid \mathcal{G}_{k-1}^{\mathcal{O}} \right) \\ & \leq \frac{1}{|\mathbb{T}_n^*|^{1/2}} \left(\sum_{k=1}^{\nu_n} \frac{\mathbb{E}[\|D_k\|^4 | \mathcal{G}_{k-1}^{\mathcal{O}}]}{|\mathbb{T}_n^*|} \right)^{1/2} \left(\sum_{k=1}^{\nu_n} \frac{\mathbb{E}[\|D_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}]}{\epsilon^2 |\mathbb{T}_n^*|} \right)^{1/2}. \end{aligned}$$

We can easily prove that $\mathbb{E}[\|D_k\|^4 | \mathcal{G}_{k-1}^{\mathcal{O}}]$ and $\mathbb{E}[\|D_k\|^2 | \mathcal{G}_{k-1}^{\mathcal{O}}]$ are polynomials functions in X_k of degree 8 and 4. Thanks to $\gamma \geq \kappa \geq 8$ and Proposition 4.11, we get the Lindeberg condition. We can now conclude that under $\mathbb{P}_{\bar{\mathcal{E}}}$ one has

$$\frac{1}{|\mathbb{T}_{n-1}^*|^{1/2}} \sum_{k \in \mathbb{T}_{n-1}^*} D_k = \frac{1}{|\mathbb{T}_{n-1}^*|^{1/2}} \mathbf{M}_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}).$$

Finally, result (3.5) follows from Eq. (5.10) and Proposition 4.14 together with Slutsky's Lemma. \square

Proof of Theorem 3.5, second step We apply Theorem 2.1.9 of [22] again to the sequences $(\mathbf{M}_p^{\sigma(n)})_{\{p \geq 1\}}$ of $\mathcal{G}_p^{\mathcal{O}}$ -martingales defined by

$$|\mathbb{T}_{n-1}^*|^{1/2} \mathbf{M}_p^{\sigma(n)} = \sum_{k=1}^p \mathbf{D}_k^{\sigma}, \quad \mathbf{D}_k^{\sigma} = \begin{pmatrix} \epsilon_{2k}^2 + \epsilon_{2k+1}^2 - \mathbb{E}[\epsilon_{2k}^2 + \epsilon_{2k+1}^2 | \mathcal{F}_{\ell}^{\mathcal{O}}] \\ 2X_k(\epsilon_{2k}^2 - \mathbb{E}[\epsilon_{2k}^2 | \mathcal{F}_{\ell}^{\mathcal{O}}]) \\ 2X_k(\epsilon_{2k+1}^2 - \mathbb{E}[\epsilon_{2k+1}^2 | \mathcal{F}_{\ell}^{\mathcal{O}}]) \\ X_k^2(\epsilon_{2k}^2 + \epsilon_{2k+1}^2 - \mathbb{E}[\epsilon_{2k}^2 + \epsilon_{2k+1}^2 | \mathcal{F}_{\ell}^{\mathcal{O}}]) \end{pmatrix}.$$

Set $\nu_n = |\mathbb{T}_{n-1}^*| = 2^n - 1$, thus, one has $|\mathbb{T}_{n-1}^*|^{1/2} \mathbf{M}_{\nu_n}^{\sigma(n)} = \mathbf{U}_{n-1}(\boldsymbol{\sigma}_n - \boldsymbol{\sigma})$. We have to study the limit $\mathbf{\Gamma}^{\sigma}$ of $|\mathbb{T}_{n-1}^*|^{-1} \sum_{k \in \mathbb{T}_{n-1}^*} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k^{\sigma} (\mathbf{D}_k^{\sigma})^t | \mathcal{G}_{k-1}^{\mathcal{O}}]$. In order to compute the conditional expectation, recall that $\mathbb{E}[\varepsilon_2^p \eta_2^q \varepsilon_3^r \eta_3^s] = \vartheta(p, q, r, s)$, and let us denote, for $k \geq 1$

$$\begin{aligned} A_i(k) &= \delta_{2k+i} \left(\sum_{r=0}^4 C_4^r \vartheta((1-i)(4-r), (1-i)r, i(4-r), ir) X_k^r \right. \\ & \quad \left. - (\sigma_{\varepsilon}^4 + 4\rho_{ii}\sigma_{\varepsilon}^2 X_k + (4\rho_{ii}^2 + 2\sigma_{\varepsilon}^2 \sigma_{\eta}^2) X_k^2 + 4\rho_{ii}\sigma_{\eta}^2 X_k^3 + \sigma_{\eta}^4 X_k^4) \right), \\ A_{01}(k) &= \delta_{2k} \delta_{2k+1} \left(\sum_{r=0}^2 \sum_{s=0}^2 C_2^r C_2^s \vartheta(2-r, r, 2-s, s) X_k^{r+s} \right. \\ & \quad \left. - (\sigma_{\varepsilon}^4 + 2\sigma_{\varepsilon}^2(\rho_{00} + \rho_{11}) X_k + (2\sigma_{\varepsilon}^2 \sigma_{\eta}^2 + 4\rho_{00}\rho_{11}) X_k^2 \right. \\ & \quad \left. + 2\sigma_{\eta}^2(\rho_{00} + \rho_{11}) X_k^3 + \sigma_{\eta}^4 X_k^4) \right), \end{aligned}$$

and $B_i(k) = A_i(k) + A_{01}(k)$. Using these notations, we obtain

$$\begin{aligned} & \mathbb{E}_{\overline{\mathcal{F}}}[\mathbf{D}_k^\sigma (\mathbf{D}_k^\sigma)^t \mid \mathcal{G}_{k-1}^{\mathcal{O}}] \\ &= \begin{pmatrix} (B_0 + B_1)(k) & 2X_k B_0(k) & 2X_k B_1(k) & X_k^2 (B_0 + B_1)(k) \\ 2X_k B_0(k) & 4X_k^2 A_0(k) & 4X_k^2 A_{01}(k) & 2X_k^3 B_0(k) \\ 2X_k B_1(k) & 4X_k^2 A_{01}(k) & 4X_k^2 A_1(k) & 2X_k^3 B_1(k) \\ X_k^2 (B_0 + B_1)(k) & 2X_k^3 B_0(k) & 2X_k^3 B_1(k) & X_k^4 (B_0 + B_1)(k) \end{pmatrix}. \end{aligned}$$

We obtain the $\mathbb{P}_{\overline{\mathcal{F}}}$ a.s. limit of the above quantity thanks to Proposition 4.11:

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}} A_i(k) X_k^q = A_i^q \quad \text{and} \quad \lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}} A_{01}(k) X_k^q = A_{01}^q,$$

with

$$\begin{aligned} A_i^q &= \sum_{r=0}^4 C_4^r \vartheta((1-i)(4-r), (1-i)r, i(4-r), ir) \ell_i(r+q) \\ &\quad - (\sigma_\varepsilon^4 \ell_i(q) + 4\rho_{ii} \sigma_\varepsilon^2 \ell_i(1+q) + (4\rho_{ii}^2 + 2\sigma_\varepsilon^2 \sigma_\eta^2) \ell_i(2+q) \\ &\quad + 4\rho_{ii} \sigma_\eta^2 \ell_i(3+q) + \sigma_\eta^4 \ell_i(4+q)), \\ A_{01}^q &= \sum_{r=0}^2 \sum_{s=0}^2 C_2^r C_2^s \vartheta(2-r, r, 2-s, s) \ell_{01}(r+s+q) \\ &\quad - (\sigma_\varepsilon^4 \ell_{01}(q) + 2\sigma_\varepsilon^2 (\rho_{00} + \rho_{11}) \ell_{01}(1+q) + (2\sigma_\varepsilon^2 \sigma_\eta^2 + 4\rho_{00} \rho_{11}) \ell_{01}(2+q) \\ &\quad + 2\sigma_\eta^2 (\rho_{00} + \rho_{11}) \ell_{01}(3+q) + \sigma_\eta^4 \ell_{01}(4+q)). \end{aligned}$$

We also set $B_i^q = A_i^q + A_{01}^q$. With these notations, we are able to explicit the limit matrix $\mathbf{\Gamma}^\sigma$ of $\langle \mathbf{M}^{\sigma(n)} \rangle_{\nu_n}$

$$\mathbf{\Gamma}^\sigma = \begin{pmatrix} B_0^0 + B_1^0 & 2B_0^1 & 2B_1^1 & B_0^2 + B_1^2 \\ 2B_0^1 & 4A_0^2 & 4A_{01}^2 & 2B_0^3 \\ 2B_1^1 & 4A_{01}^2 & 4A_1^2 & 2B_1^3 \\ B_0^2 + B_1^2 & 2B_0^3 & 2B_1^3 & B_0^4 + B_1^4 \end{pmatrix} \quad (6.1)$$

The first assumption of Theorem 2.1.9 of [22] holds under $\mathbb{P}_{\overline{\mathcal{F}}}$ and we prove the second one as in the first step. We then conclude that under $\mathbb{P}_{\overline{\mathcal{F}}}$ one has

$$\mathbf{M}_{\nu_n}^{\sigma(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k \in \mathbb{T}_{n-1}} \mathbf{D}_k^\sigma = |\mathbb{T}_{n-1}^*|^{-1/2} \mathbf{U}_{n-1} (\boldsymbol{\sigma}_n - \boldsymbol{\sigma}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}^\sigma).$$

We conclude using Proposition 4.14, Theorem 3.3 and Slutsky's Lemma. \square

Proof of Theorem 3.5, third step We use again Theorem 2.1.9 of [22] with to the sequence of $\mathcal{G}_p^{\mathcal{O}}$ -martingales $(\mathbf{M}_p^{\rho(n)})_{\{p \geq 1\}}$ defined by

$$|\mathbb{T}_{n-1}^*|^{1/2} \mathbf{M}_p^{\rho(n)} = \sum_{k=1}^p \mathbf{D}_k^\rho, \quad \mathbf{D}_k^\rho = \begin{pmatrix} \epsilon_{2k} \epsilon_{2k+1} - \mathbb{E}[\epsilon_{2k} \epsilon_{2k+1} \mid \mathcal{F}_\ell^{\mathcal{O}}] \\ 2X_k (\epsilon_{2k} \epsilon_{2k+1} - \mathbb{E}[\epsilon_{2k} \epsilon_{2k+1} \mid \mathcal{F}_\ell^{\mathcal{O}}]) \\ X_k^2 (\epsilon_{2k} \epsilon_{2k+1} - \mathbb{E}[\epsilon_{2k} \epsilon_{2k+1} \mid \mathcal{F}_\ell^{\mathcal{O}}]) \end{pmatrix}.$$

Setting again $\nu_n = |\mathbb{T}_{n-1}|$, we have $|\mathbb{T}_{n-1}^*|^{1/2} \mathbf{M}_{\nu_n}^{\rho(n)} = \mathbf{V}_{n-1}(\boldsymbol{\rho}_n - \boldsymbol{\rho})$. Set

$$\begin{aligned} C(k) &= \delta_{2k} \delta_{2k+1} \left((\vartheta(2, 0, 2, 0) - \rho_\varepsilon^2) + 2(\vartheta(2, 0, 1, 1) + \vartheta(1, 1, 2, 0) - 2\rho\rho_\varepsilon) X_k \right. \\ &\quad + (\vartheta(0, 2, 2, 0) + \vartheta(2, 0, 0, 2) + 4\vartheta(1, 1, 1, 1) - 4\rho^2 - 2\rho_\varepsilon \rho_\eta) X_k^2 \\ &\quad \left. + 2(\vartheta(0, 2, 1, 1) + \vartheta(1, 1, 0, 2) - 2\rho\rho_\eta) X_k^3 + (\vartheta(0, 2, 0, 2) - \rho_\eta^2) X_k^4 \right), \end{aligned}$$

so that we are now able to write

$$\mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k^\rho (\mathbf{D}_k^\rho)^t \mid \mathcal{G}_{k-1}^\mathcal{O}] = C(k) \begin{pmatrix} 1 & 2X_k & X_k^2 \\ 2X_k & 4X_k^2 & 2X_k^3 \\ X_k^2 & 2X_k^3 & X_k^4 \end{pmatrix}.$$

For the determination of the limit $\boldsymbol{\Gamma}^\rho$ of $|\mathbb{T}_{n-1}^*|^{-1} \sum_{k \in \mathbb{T}_{n-1}} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k^\rho (\mathbf{D}_k^\rho)^t \mid \mathcal{G}_{k-1}^\mathcal{O}]$, let us remark, using Proposition 4.11, that

$$\lim_{n \rightarrow \infty} \frac{1}{|\mathbb{T}_{n-1}^*|} \sum_{k \in \mathbb{T}_{n-1}} C(k) X_k^q = C^q \quad \text{a.s.}$$

with

$$\begin{aligned} C^q &= (\vartheta(2, 0, 2, 0) - \rho_\varepsilon^2) \ell_{01}(q) + 2(\vartheta(2, 0, 1, 1) + \vartheta(1, 1, 2, 0) - 2\rho\rho_\varepsilon) \ell_{01}(1+q) \\ &\quad + (\vartheta(0, 2, 2, 0) + \vartheta(2, 0, 0, 2) + 4\vartheta(1, 1, 1, 1) - 4\rho^2 - 2\rho_\varepsilon \rho_\eta) \ell_{01}(2+q) \\ &\quad + 2(\vartheta(0, 2, 1, 1) + \vartheta(1, 1, 0, 2) - 2\rho\rho_\eta) \ell_{01}(3+q) \\ &\quad + (\vartheta(0, 2, 0, 2) - \rho_\eta^2) \ell_{01}(4+q). \end{aligned}$$

The matrix $\boldsymbol{\Gamma}^\rho$ is thus given by

$$\boldsymbol{\Gamma}^\rho = \begin{pmatrix} C^0 & 2C^1 & C^2 \\ 2C^1 & 4C^2 & 2C^3 \\ C^2 & 2C^3 & C^4 \end{pmatrix}. \quad (6.2)$$

The first assumption of Theorem 2.1.9 of [22] holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. We prove the second one as in the previous steps, and finally obtain that under $\mathbb{P}_{\bar{\mathcal{E}}}$ one has

$$\mathbf{M}_{\nu_n}^{\rho(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k \in \mathbb{T}_{n-1}} \mathbf{D}_k^\rho = |\mathbb{T}_{n-1}^*|^{-1/2} \mathbf{V}_{n-1}(\boldsymbol{\rho}_n - \boldsymbol{\rho}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{\Gamma}^\rho).$$

We conclude using Proposition 4.14 and Theorem 3.4. \square

7 Application to real data

We have applied our estimation procedure to the *Escherichia coli* data of [11]. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell has thus a new end (or *pole*), and an older one. The cell inheriting the old pole of its mother is called the old pole cell, its sister is called the new pole cell. Hence,

a	b	c	d
0.0363	0.0266	0.0306	0.1706
[0.0275, 0.0450]	[-0.2094, 0.2627]	[0.0216, 0.0396]	[-0.0709, 0.4120]

Table 1: Estimation of $\boldsymbol{\theta}$ on the data set *penna-2002-10-04-4*

σ_ε^2	σ_η^2
0.0004	0.2431
[-0, 0002, 0.0010]	[-0.0750, 0.5613]

Table 2: Estimation of noise variances on the data set *penna-2002-10-04-4*

each cell has a *type*: old pole (even) or new pole (odd), inducing asymmetry in the cell division. Stewart et al. [11] filmed colonies of dividing cells, determining the complete lineages and the growth rate of each cell. Several attempts have already been made to fit BAR processes to these data, see [20, 12, 15, 16], but only with fixed coefficients models. In particular, [20] suggests that such models cannot explain all the randomness of the data. We have run our estimators on the data set *penna-2002-10-04-4*. It is the largest one of the experiment. It contains 663 cells up to generation 9. Table 1 gives the estimation $\widehat{\boldsymbol{\theta}}_9$ of $\boldsymbol{\theta}$ with the 95% Confidence Interval (C.I.) of each coefficient. Note that our estimator $\widehat{\boldsymbol{\theta}}_n$ is exactly the same as in [15], so that we obtain the same point estimation. The confidence intervals are wider, as the variance is different. More precisely, the variance is given by the CLT for $\boldsymbol{\theta}$ in Eq. (3.5). It can be approximated by $|\mathbb{T}_8^*| \mathbf{S}_8^{-1} \boldsymbol{\Gamma}_8 \mathbf{S}_8^{-1}$ thanks to the convergences given in Proposition 4.14 and Lemma 5.4. Table 2 gives the estimation of the variance coefficients σ_ε^2 and σ_η^2 of $\boldsymbol{\sigma}$ (other covariance coefficients of $\widehat{\boldsymbol{\sigma}}_9$ and $\widehat{\boldsymbol{\rho}}_9$ can be computed but are less easy to interpret). The variance of these parameters is again given by the central limit Theorem 3.5. To obtain confidence intervals, one needs an estimation of the joint moments of $(\varepsilon_2, \eta_2, \varepsilon_3, \eta_3)$ up to the order 4. Such estimators can be easily derived following the same ideas as in Section 3.1. Theorem 3.5 allows to build a positivity test for σ_ε^2 and σ_η^2 . The p-value of the test $H_0 : \sigma_\varepsilon^2 = 0$ (resp. $H_0 : \sigma_\eta^2 = 0$) is $p = 0.0799$ (resp. $p = 0.0671$). We are not far to support the validity of the random coefficients model on this data set.

References

- [1] R. Cowan, R. G. Staudte, The bifurcating autoregressive model in cell lineage studies, *Biometrics* 42 (1986) 769–783.
- [2] R. M. Huggins, Robust inference for variance components models for single trees of cell lineage data, *Ann. Statist.* 24 (3) (1996) 1145–1160.
- [3] R. M. Huggins, I. V. Basawa, Inference for the extended bifurcating autoregressive model for cell lineage studies, *Aust. N. Z. J. Stat.* 42 (4) (2000) 423–432.

- [4] J. Zhou, I. V. Basawa, Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors, *J. Time Ser. Anal.* 26 (6) (2005) 825–842.
- [5] R. M. Huggins, R. G. Staudte, Variance components models for dependent cell populations, *J. AMS* 89 (425) (1994) 19–29.
- [6] R. M. Huggins, I. V. Basawa, Extensions of the bifurcating autoregressive model for cell lineage studies, *J. Appl. Probab.* 36 (4) (1999) 1225–1233.
- [7] S. Y. Hwang, I. V. Basawa, Asymptotic optimal inference for multivariate branching-Markov processes via martingale estimating functions and mixed normality, *J. Multivariate Anal.* 102 (6) (2011) 1018–1031.
- [8] S. Y. Hwang, I. V. Basawa, Branching Markov processes and related asymptotics, *J. Multivariate Anal.* 100 (6) (2009) 1155–1167.
- [9] I. V. Basawa, J. Zhou, Non-Gaussian bifurcating models and quasi-likelihood estimation, *J. Appl. Probab.* 41A (2004) 55–64.
- [10] J. Zhou, I. V. Basawa, Least-squares estimation for bifurcating autoregressive processes, *Statist. Probab. Lett.* 74 (1) (2005) 77–88.
- [11] E. Stewart, R. Madden, G. Paul, F. Taddei, Aging and death in an organism that reproduces by morphologically symmetric division, *PLoS Biol.* 3 (2) (2005) e45.
- [12] J. Guyon, Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging, *Ann. Appl. Probab.* 17 (5-6) (2007) 1538–1569.
- [13] B. Bercu, B. de Saporta, A. Gégout-Petit, Asymptotic analysis for bifurcating autoregressive processes via a martingale approach, *Electron. J. Probab.* 14 (2009) no. 87, 2492–2526.
- [14] J.-F. Delmas, L. Marsalle, Detection of cellular aging in a Galton-Watson process, *Stoch. Process. and Appl.* 120 (2010) 2495–2519.
- [15] B. de Saporta, A. Gégout-Petit, L. Marsalle, Parameters estimation for asymmetric bifurcating autoregressive processes with missing data, *Electron. J. Statist.* 5 (2011) 1313–1353.
- [16] B. de Saporta, A. Gégout Petit, L. Marsalle, Asymmetry tests for bifurcating autoregressive processes with missing data, *Statistics & Probability Letters* 82 (7) (2012) 1439–1444.
- [17] V. Bansaye, Proliferating parasites in dividing cells: Kimmel’s branching model revisited, *Ann. Appl. Probab.* 18 (3) (2008) 967–996.
- [18] Q. M. Bui, R. M. Huggins, Inference for the random coefficients bifurcating autoregressive model for cell lineage studies, *J. Statist. Plann. Inference* 81 (2) (1999) 253–262.

- [19] V. Blandin, Asymptotic results for bifurcating random coefficient autoregressive processes, arXiv 1204.2926 (2012).
- [20] J. Guyon, A. Bize, G. Paul, E. Stewart, J.-F. Delmas, F. Taddéi, Statistical study of cellular aging, in: CEMRACS 2004—mathematics and applications to biology and medicine, Vol. 14 of ESAIM Proc., EDP Sci., Les Ulis, 2005, pp. 100–114 (electronic).
- [21] D. F. Nicholls, B. G. Quinn, Random coefficient autoregressive models: an introduction, Vol. 11 of Lecture Notes in Statistics, Springer-Verlag, New York, 1982, lecture Notes in Physics, 151.
- [22] M. Duflo, Random iterative models, Vol. 34 of Applications of Mathematics, Springer-Verlag, Berlin, 1997.
- [23] P. Hall, C. C. Heyde, Martingale limit theory and its application, Academic Press Inc., New York, 1980, probability and Mathematical Statistics.
- [24] J. D. Hamilton, Time series analysis, Princeton University Press, Princeton, NJ, 1994.
- [25] A. Brandt, The stochastic equation $Y_{n+1} = A_n Y_n + B_n$ with stationary coefficients, Adv. in Appl. Probab. 18 (1) (1986) 211–220.
- [26] B. de Saporta, Tail of the stationary solution of the stochastic equation $Y_{n+1} = a_n Y_n + b_n$ with Markovian coefficients, Stochastic Process. Appl. 115 (12) (2005) 1954–1978.
- [27] C. Z. Wei, Adaptive prediction by least squares predictors in stochastic regression models with applications to time series, Ann. Statist. 15 (4) (1987) 1667–1682.

- 1.5 [dSGP⁺13b] B. DE SAPORTA, A. GÉGOUT PETIT, AND L. MARSALLE. Statistical study of asymmetry in cell lineage data. En révision pour *Comp.l Statistics & Data Analysis*, arXiv :1205.4840, 2013.

Statistical study of asymmetry in cell lineage data

Benoîte de Saporta

Univ. Bordeaux, Gretha, UMR 5113, IMB, UMR 5251, F-33400 Talence, France
CNRS, Gretha, UMR 5113, IMB, UMR 5251, F-33400 Talence, France
INRIA Bordeaux Sud Ouest, team CQFD, F-33400 Talence, France

Anne Gégout-Petit

Univ. Bordeaux, IMB, UMR 5251, F-33400 Talence, France
CNRS, IMB, UMR 5251, F-33400 Talence, France
INRIA Bordeaux Sud Ouest, team CQFD, F-33400 Talence, France

Laurence Marsalle

Univ. Lille 1, Laboratoire Paul Painlevé, UMR 8524, F-59 655 Villeneuve d'Ascq, France
CNRS, Laboratoire Paul Painlevé, UMR 8524, F-59 655 Villeneuve d'Ascq, France

March 11, 2013

Abstract

This paper proposes a rigorous methodology to study cell division data consisting in several observed genealogical trees of possibly different shapes. Our procedure allows us to fully take into account missing observations, data from different trees as well as the dependence structure within genealogical trees. It also enables us to use all the information available without the drawbacks of low accuracy for estimators or low power for tests on small single trees. We model the data by an asymmetric bifurcating autoregressive process and take into account possibly missing observations by modeling the genealogies with a two-type Galton Watson process. We give least-squares estimators of the unknown parameters of the processes and derive symmetry tests. Our results are applied on real data of *Escherichia coli* division.

1 Introduction

Cell lineage data consist of observations of some quantitative characteristic of the cells (e.g. their length, growth rate, ...) over several generations descended from an initial cell. Track is kept of the genealogy to study the inherited effects on the evolution of the characteristic. As a cell usually gives birth to two offspring by division, such genealogies are structured as binary trees. [2] first adapted autoregressive processes to this binary tree structure by introducing bifurcating autoregressive processes (BAR). This parametric model takes into account both the environmental and inherited effects. Inference on this model has been proposed based on either a single tree growing to infinity, see e.g. [2], [11], [13], [17] or for an asymptotically infinite number of small replicated trees, see e.g. [14], [12].

More recently, studies of aging in single cell organisms by [16] suggested that cell division may not be symmetric. An asymmetric BAR model was therefore proposed by [8], where the two sets of parameters corresponding to sister cells are allowed to be different. Inference for this model was only investigated for single trees growing to infinity, see [8], [1] for the fully observed model or [5], [3], [4] where missing data are taken into account.

Cell division data often consist in recordings over several genealogies of cells evolving in similar experimental conditions. For instance, [16] filmed 94 colonies of *Escherichia coli* cells dividing between four and nine times. We therefore propose a new rigorous approach to take into account all the available information. Indeed, we propose an inference based on a finite fixed number of replicated trees when the total number of observed cells tends to infinity. We use the missing

data asymmetric BAR model introduced by [3]. In this approach, the observed genealogies are modeled with a two-type Galton Watson (GW) process. However, we propose a different least-squares estimator for the parameters of the BAR process that does not correspond to the single-tree estimators averaged on the replicated trees. We also propose an estimator of the parameters of the GW process specific to our binary tree structure and not based simply on the observation of the number of cells of each type in each generation as in [7], [15]. We study the consistency and asymptotic normality of our estimators and derive asymptotic confidence intervals as well as Wald's type tests to investigate the asymmetry of the data for both the BAR and GW processes. Our results are applied to the *Escherichia coli* data of [16].

The paper is organized as follows. In Section 2, we present the BAR and observations models. In Section 3 we give our estimators and state their asymptotic properties. In Section 4, we propose a new investigation of [16] data. The precise statement of the convergence results, the explicit form of the asymptotic variance of the estimators and the convergence proofs are postponed to the appendix.

2 Model

Our aim is to estimate the parameters of coupled BAR and GW processes through m i.i.d. realizations of the processes. We first define our parametric model and introduce our notation. The BAR and GW processes have the same dynamics as in [3], the main difference is that our inference is here based on several i.i.d. realizations of the processes, instead of a single one. Additional notation together with the precise technical assumptions are specified in A.

2.1 Bifurcating autoregressive model

Our model is as follows. Consider m i.i.d. replications of the asymmetric BAR process with coefficients $(a, b, c, d) \in \mathbb{R}^4$. More precisely, for $1 \leq j \leq m$, the first cell in genealogy j is labelled $(j, 1)$ and for $k \geq 1$, the two offspring of cell (j, k) are labelled $(j, 2k)$ and $(j, 2k+1)$. As we consider an asymmetric model, each cell has a *type* defined by its label: $(j, 2k)$ has type *even* and $(j, 2k+1)$ has type *odd*. The characteristic of cell k in genealogy j is denoted by $X_{(j,k)}$. The BAR processes are defined recursively as follows: for all $1 \leq j \leq m$ and $k \geq 1$, one has

$$\begin{cases} X_{(j,2k)} &= a + bX_{(j,k)} + \varepsilon_{(j,2k)}, \\ X_{(j,2k+1)} &= c + dX_{(j,k)} + \varepsilon_{(j,2k+1)}. \end{cases} \quad (1)$$

Let also σ^2 be the variance and ρ the covariance of the noise sequence

$$\sigma^2 = \mathbb{E}[\varepsilon_{(j,2k)}^2] = \mathbb{E}[\varepsilon_{(j,2k+1)}^2], \quad \rho = \mathbb{E}[\varepsilon_{(j,2k)}\varepsilon_{(j,2k+1)}].$$

Our goal is to estimate the parameters (a, b, c, d) and (σ, ρ) , and then test if $(a, b) = (c, d)$ or not.

2.2 Observation process

We now turn to the observation process $(\delta_{(j,k)})$ that encodes for the presence or absence of cell measurements in the available data

$$\delta_{(j,k)} = \begin{cases} 1 & \text{if cell } k \text{ in genealogy } j \text{ is observed,} \\ 0 & \text{if cell } k \text{ in genealogy } j \text{ is not observed.} \end{cases}$$

To take into account possible asymmetry in the observation process, we use a two-type Galton Watson model. The relevance of this model to *E. coli* data is discussed in section 4. Here again we suppose all the m observation processes to be drawn independently from the same two-type GW process. More precisely, for all $1 \leq j \leq m$, we define the observation process $(\delta_{(j,k)})_{k \geq 1}$ for the j -th genealogy as follows. We set $\delta_{(j,1)} = 1$ and the $(\delta_{(j,2k)}, \delta_{(j,2k+1)})$ are drawn independently from one another with a law depending on the type of cell k . More precisely, for $i \in \{0, 1\}$, if k is of type i one has

$$\mathbb{P}\left((\delta_{(j,2k)}, \delta_{(j,2k+1)}) = (l_0, l_1) \mid \delta_{(j,k)} = 1\right) = p^{(i)}(l_0, l_1),$$

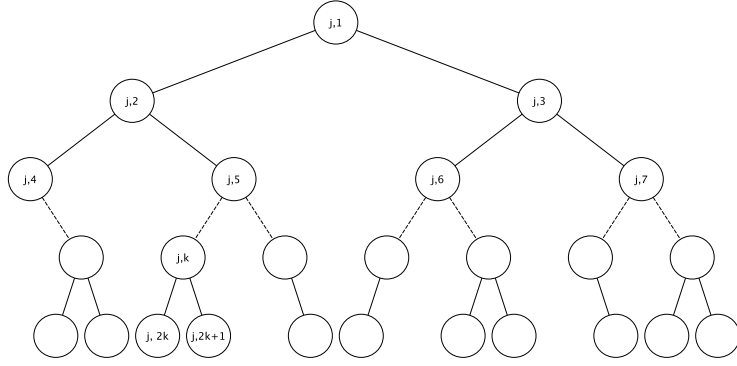


Figure 1: A tree of observed cells.

for all $(l_0, l_1) \in \{0, 1\}^2$. Thus, $p^{(i)}(l_0, l_1)$ is the probability that a cell of type i has l_0 daughter of type 0 and l_1 daughter of type 1. If a cell is not observed, all its descendants are missing as well: if $\delta_{(j,k)} = 0$ for some $k \geq 2$, then for all its descendants l , $\delta_{(j,l)} = 0$. Figure 1 gives an example of realization of an observation process. We also assume that the observation processes are independent from the BAR processes.

3 Inference

Our first goal is to estimate the reproduction probabilities $p^{(i)}(l_0, l_1)$ of the GW process from the m genealogies of observed cells up to the n -th generation to be able to test the symmetry of the GW model itself. Our second goal is to estimate $\theta = (a, b, c, d)^t$ from all the observed individuals of the m trees up to the n -th generation. We then give the asymptotic properties of our estimator to be able to build confidence intervals and symmetry tests for θ .

Denote by $|\mathbb{T}_n^*|$ the total number of observed cells in the m trees up to the n -th generation of offspring from the original ancestors, and let

$$\bar{\mathcal{E}} = \left\{ \lim_{n \rightarrow \infty} |\mathbb{T}_n^*| = \infty \right\}$$

be the non extinction set, on which the global cell population grows to infinity. Conditions for the probability of non-extinction to be positive are given in A.

3.1 Estimation of the reproduction law of the GW process

There are many references on the inference of a multi-type GW process, see for instance [7] and [15]. Our context of estimation is very specific because the information given by $(\delta_{(j,k)})$ is more precise than that given by the number of cells of each type in a given generation that is usually used in the literature. Indeed, not only do we know the number of cells of each type at each generation, but we also know their precise positions on the binary tree of cell division. The empiric estimators of the reproduction probabilities using data up to the n -th generation are then, for i, l_0, l_1 in $\{0, 1\}$

$$\hat{p}_n^{(i)}(l_0, l_1) = \frac{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+i)} \phi_{l_0}(\delta_{(j,2(2k+i))}) \phi_{l_1}(\delta_{(j,2(2k+i)+1})}{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+i)}}$$

where $\phi_0(x) = 1 - x$, $\phi_1(x) = x$, and if the denominator is non zero, the estimator equalling zero otherwise. Note that the numerator is just the number of cells of type i in all the trees up to generation $n - 1$ that have exactly l_0 daughter of type 0 and l_1 daughter of type 1 in the n -th generation. The denominator is the total number of cells of type i in all the trees up to generation $n - 1$. Set also

$$\mathbf{p}^{(i)} = (p^{(i)}(0, 0), p^{(i)}(0, 1), p^{(i)}(1, 0), p^{(i)}(1, 1))^t$$

the vector of the 4 reproduction probabilities for a mother of type i , $\mathbf{p} = ((\mathbf{p}^{(0)})^t, (\mathbf{p}^{(1)})^t)^t$ the vector of all 8 reproduction probabilities and $\hat{\mathbf{p}}_n$ its empirical estimator.

3.2 Least-squares estimation for the BAR parameters

For the parameters of the BAR process, let us denote $\theta = (a, b, c, d)^t$. We use the standard least squares (LS) estimator $\hat{\theta}_n$ which minimizes

$$\Delta_n(\theta) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k)} (X_{(j,2k)} - a - bX_{(j,k)})^2 + \delta_{(j,2k+1)} (X_{(j,2k+1)} - c - dX_{(j,k)})^2.$$

Consequently, for all $n \geq 1$ we have $\hat{\theta}_n = (\hat{a}_n, \hat{b}_n, \hat{c}_n, \hat{d}_n)^t$ with

$$\hat{\theta}_n = \Sigma_{n-1}^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} (\delta_{(j,2k)} X_{(j,2k)}, \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)}, \delta_{(j,2k+1)} X_{(j,2k+1)}, \delta_{(j,2k+1)} X_{(j,k)} X_{(j,2k+1)})^t \quad (2)$$

where, for $i \in \{0, 1\}$ one has

$$\Sigma_n = \begin{pmatrix} \mathbf{S}_n^0 & 0 \\ 0 & \mathbf{S}_n^1 \end{pmatrix}, \quad \mathbf{S}_n^i = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j,2k+i)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}.$$

Note that in the normalizing matrices \mathbf{S}_n^i the sum is over all observed cells for which a daughter of type i is observed, and not merely over all observed cells. To estimate the variance parameters σ^2 and ρ , we define the empiric residuals. For all $2^\ell \leq k \leq 2^{\ell+1} - 1$ and $1 \leq j \leq m$ set

$$\begin{cases} \hat{\varepsilon}_{(j,2k)} & = & \delta_{(j,2k)} (X_{(j,2k)} - \hat{a}_\ell - \hat{b}_\ell X_{(j,k)}), \\ \hat{\varepsilon}_{(j,2k+1)} & = & \delta_{(j,2k+1)} (X_{(j,2k+1)} - \hat{c}_\ell - \hat{d}_\ell X_{(j,k)}). \end{cases} \quad (3)$$

We propose the following empirical estimators

$$\hat{\sigma}_n^2 = \frac{1}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} (\hat{\varepsilon}_{(j,2k)}^2 + \hat{\varepsilon}_{(j,2k+1)}^2), \quad \hat{\rho}_n = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k)} \hat{\varepsilon}_{(j,2k+1)},$$

where $|\mathbb{T}_n^{*01}|$ is the set of all the cells which have exactly two offspring in the m trees up to generation n .

3.3 Consistency and normality

We can now state the convergence results we obtain for the estimators above. The assumptions (H.1) to (H.6) are given in A.2. These results hold on the non extinction set $\bar{\mathcal{E}}$.

Theorem 3.1 *Under assumptions (H.5-6) and for all i, l_0 and l_1 in $\{0, 1\}$, $\hat{p}_n^{(i)}(l_0, l_1)$ converges to $p^{(i)}(l_0, l_1)$ almost surely on $\bar{\mathcal{E}}$. Under assumptions (H.0-6), $\hat{\theta}_n$, $\hat{\sigma}_n^2$ and $\hat{\rho}_n$ converge to θ , σ^2 and ρ respectively, almost surely on $\bar{\mathcal{E}}$.*

The asymptotic normality results are only valid conditionally to the non extinction of the global cell population.

Theorem 3.2 *Under assumptions (H.5-6) we have*

$$\sqrt{|\mathbb{T}_{n-1}^*|} (\hat{\mathbf{p}}_n - \mathbf{p}) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{V}),$$

and under assumptions (H.0-6), we have

$$\sqrt{|\mathbb{T}_{n-1}^*|} (\hat{\theta}_n - \theta) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}_\theta), \quad \sqrt{|\mathbb{T}_n^*|} (\hat{\sigma}_n^2 - \sigma^2) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma_\sigma), \quad \sqrt{|\mathbb{T}_{n-1}^{*01}|} (\hat{\rho}_n - \rho) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \gamma_\rho),$$

conditionally to $\bar{\mathcal{E}}$. The explicit form of the variance matrices \mathbf{V} and $\mathbf{\Gamma}_\theta$ and of γ_σ and γ_ρ is given in Eq. (6), (7), (9) and (10) respectively.

The proofs of these results are given in B.2 and B.3 for the GW process and in C.2 and C.3 for the BAR process. From the asymptotic normality, one can naturally construct confidence intervals and tests. Their explicit formulas are given in B.4 and C.4.

4 Data analysis

We applied our procedure to the *Escherichia coli* data of [16]. The biological issue addressed is aging in single cell organisms. *E. coli* is a rod-shaped bacterium that reproduces by dividing in the middle. Each cell inherits an old end or *pole* from its mother, and a new one coming from the division. The cell that inherits the old pole of its mother is called the *old pole* cell, the other one is called the *new pole* cell. Therefore, each cell has a *type*: old pole or new pole cell that can be interpreted as an *age*, inducing asymmetry in the cell division. On a binary tree, the new pole cells are labelled by an even number and the old pole cells by an odd number.

[16] filmed 94 colonies of dividing *E. coli* cells, determining the complete lineage and the growth rate of each cell. The 94 data sets gather 22394 cells (11189 of type even and 11205 of type odd). The number of divisions goes from four to nine. Not a single data tree is complete. Missing data mainly do not come from cell death (only 16 cells are recorded to die) but from measurement difficulties due mostly to overlapping cells or cells wandering away from the field of view. Note also that for a growth rate to be recorded, the cell needs to be observed through its whole life cycle. If this is not the case, there is no record at all, so that a censored data model is not relevant. The observed average growth rate of even (resp. odd) cells is 0.0371 (resp. 0.0369). These data were investigated in [16, 9, 8, 3, 4].

[16] proposed a statistical study of the averaged genealogy and pair-wise comparison of sister cells. They concluded that the old pole cells exhibit cumulatively slowed growth, less offspring biomass production and an increased probability of death whereas single-experiment analyses did not. However they assumed independence between the averaged couples of sister cells, which does not hold in such genealogies.

The other studies are based on single-tree analyses instead of averaging all the genealogical trees. [9] model the growth rate by a Markovian bifurcating process, but their procedure does not take into account the dependence between pairs of sister cells either. The asymmetry was rejected (p -value < 0.1) in half of the experiments so that a global conclusion was difficult. [8] has then investigated the asymptotical properties of a more general asymmetric Markovian bifurcating autoregressive process, and he rigorously constructed a Wald's type test to study the asymmetry of the process. However, his model does not take into account the possible missing data from the genealogies. The author investigates the method on the 94 data sets but it is not clear how he manages missing data. More recently, [3] proposed a single-tree analysis with a rigorous method to deal with the missing data and carried out their analysis on the largest data set, concluding to asymmetry on this single set. Further single tree studies of the 51 data sets issued from the 94 colonies containing at least 8 generations were conducted in [4]. The symmetry hypothesis is rejected in one set out of four for $(a, b) = (c, d)$ and one out of eight for $a/(1 - b) = c/(1 - d)$ forbidding a global conclusion. Simulation studies tend to prove that the power of the tests on single trees is quite low for only eight or nine generations. This is what motivated the present study and urged us to use all the data available in one global estimation rather than single tree analyses.

In this section, we propose a new investigation of *E. coli* data of [16] where for the first time the dependence structure between cells within a genealogy is fully taken into account, missing data are taken care of rigorously, all the available data, i.e. the 94 sets, are analyzed at once and both the growth rate and the number/type of descendants are investigated. It is sensible to consider that all the data sets correspond to BAR processes with the same coefficients as the experiments where conducted in similar conditions. Moreover, a direct comparison of single tree estimations would be meaningless as the data trees do not all have the same number of generations, and it would be impossible to determine whether variations in the computed single tree estimators come from an intrinsic variability between trees or just the low accuracy of the estimators for small trees. The original estimation procedure described in this paper enables us to use all the information available without the drawbacks of low accuracy for estimators or low power for tests on small single trees.

4.1 Symmetry of the BAR process

We now give the results of our new investigation of the *E. coli* growth rate data of [16]. We suppose that the growth rate of cells in each lineage is modeled by the BAR process defined in

Eq. (1) and observed through the two-type GW process defined in section 2.2. The experiments were independent and lead in the same conditions corresponding to independence and identical distribution of the processes $(X_{(j,\cdot)}, \delta_{(j,\cdot)})$, $1 \leq j \leq m$.

We first give the point and interval estimation for the various parameters of the BAR process. Table 1 gives the estimation $\hat{\theta}_9$ of θ with the 95% confidence interval (CI) of each coefficient. The

Table 1: Estimation and 95 % CI of θ .

parameter	estimation	CI	parameter	estimation	CI
a	0.0203	[0.0197; 0.0210]	c	0.0195	[0.0188; 0.0201]
b	0.4615	[0.4437; 0.4792]	d	0.4782	[0.4605; 0.4959]

confidence intervals of b and d show that the *non explosion* assumption $|b| < 1$ and $|d| < 1$ is satisfied.

Table 2 gives the estimation $\hat{\sigma}_9^2$ of σ^2 and $\hat{\rho}_9$ of ρ with the 95% CI of each coefficient.

Table 2: Estimation and 95 % CI of σ^2 and ρ

parameter	estimation	CI
σ^2	$1.81 \cdot 10^{-5}$	$[1.12 \cdot 10^{-5}; 2.50 \cdot 10^{-5}]$
ρ	$0.48 \cdot 10^{-5}$	$[0.44 \cdot 10^{-5}; 0.52 \cdot 10^{-5}]$

We now turn to the results of symmetry tests. The hypothesis of equality of the couples $(a, b) = (c, d)$ is strongly rejected (p-value = 10^{-5}). The hypothesis of the equality of the two fixed points $a/(1-b)$ (estimated at 0.03773) and $c/(1-d)$ (estimated at 0.03734) of the BAR process is also rejected (p-value = $2 \cdot 10^{-3}$). We can therefore rigorously confirm that there is a statistically significant asymmetry in the division of E. coli.

4.2 Symmetry of the GW process

Let us now turn to the asymmetry of the GW process itself. Note that to our best knowledge, it is the first time this question is investigated for the E. coli data of [16]. However, it seems natural to investigate whether old pole cells have a different reproduction law from new pole ones, since aging may also induce changes in the reproduction laws.

We estimated the parameters $p^{(i)}(l_0, l_1)$ of the reproduction laws of the underlying GW process. Table 3 gives the estimations $\hat{p}_9^{(i)}(l_0, l_1)$ of the $p^{(i)}(l_0, l_1)$. The estimation of the dominant eigenvalue

Table 3: Estimation and 95 % CI of \mathbf{p} .

parameter	estimation	CI	parameter	estimation	CI
$p^{(0)}(1, 1)$	0.56060	[0.56055; 0.56065]	$p^{(1)}(1, 1)$	0.55928	[0.55923; 0.55933]
$p^{(0)}(1, 0)$	0.03621	[0.03620; 0.03622]	$p^{(1)}(1, 0)$	0.04707	[0.04706; 0.04708]
$p^{(0)}(0, 1)$	0.04740	[0.04739; 0.04741]	$p^{(1)}(0, 1)$	0.03755	[0.03754; 0.03756]
$p^{(0)}(0, 0)$	0.35579	[0.35574; 0.35583]	$p^{(1)}(0, 0)$	0.35611	[0.35606; 0.35616]

π of the descendants matrix of the GW processes (characterizing extinction, see A.1) is $\hat{\pi}_9 = 1.204$ with CI [1.191; 1.217]. The non-extinction hypothesis ($\pi > 1$) is thus satisfied.

The means of the two reproduction laws $\mathbf{p}^{(0)}$ and $\mathbf{p}^{(1)}$ are estimated at $\hat{m}_9^0 = 1.2048$ and $\hat{m}_9^1 = 1.2032$ respectively. The hypothesis of the equality of the mean numbers of offspring is not rejected (p-value = 0.9). However, Table 3 shows that there is a statistically significant difference between vectors $\mathbf{p}^{(0)}$ and $\mathbf{p}^{(1)}$ as none of the confidence intervals intersect. Indeed, the symmetry hypothesis $\mathbf{p}^{(0)} = \mathbf{p}^{(1)}$ is rejected with p-value = $2 \cdot 10^{-5}$. However, it is not possible to interpret this asymmetry in terms of the division of E. coli, since the cause of missing data is mostly due to observation difficulties rather than some intrinsic behavior of the cells.

5 Conclusion

In this paper, we first propose a statistical model to estimate and test asymmetry of a quantitative characteristic associated to each node of a family of incomplete binary trees, without aggregating single tree estimators. An immediate application is the investigation of asymmetry in cell lineage data. This model of coupled GW-BAR process generalizes all the previous methods on this subject in the literature because it rigorously takes into account:

- the dependence of the characteristic of a cell to that of its mother and the correlation between two sisters through the BAR model,
- the possibly missing data through the GW model,
- the information from several sets of data obtained in similar experimental conditions without the drawbacks of poor accuracy or power for small single trees.

Furthermore, we propose the estimation of parameters of a two-type GW process in a specific context of a binary tree and a fine observation, namely the presence or absence of each cell of the complete binary tree is known. Again the asymmetry of the parameters of the GW process could be applied to cell lineage data and be interpreted as a difference in the reproduction laws between the two different types of cell.

We applied our procedure to the *E. coli* stat of [16] and concluded there exists a statistically significant asymmetry in this cell division.

A Technical assumptions and notation

Our convergence results rely on martingale theory and the use of several carefully chosen filtrations regarding the BAR and/or GW process. The approach is similar to that of [3, 4], but their results cannot be directly applied here. This is mainly due to our choice of the global non extinction set as the union and not the intersection of the extinction sets of each replicated process. We now give some additional notation and the precise assumptions of our convergence theorems.

A.1 Generations and extinction

We first introduce some notation about the complete and observed genealogy trees that will be used in the sequel. For all $n \geq 1$, denote the n -th generation of any given tree by $\mathbb{G}_n = \{k, 2^n \leq k \leq 2^{n+1} - 1\}$. In particular, $\mathbb{G}_0 = \{1\}$ is the initial generation, and $\mathbb{G}_1 = \{2, 3\}$ is the first generation of offspring from the first ancestor. Denote by $\mathbb{T}_n = \bigcup_{\ell=0}^n \mathbb{G}_\ell$ the sub-tree of all individuals from the original individual up to the n -th generation. Note that the cardinality $|\mathbb{G}_n|$ of \mathbb{G}_n is 2^n , while that of \mathbb{T}_n is $|\mathbb{T}_n| = 2^{n+1} - 1$. Finally, we define the sets of observed individuals in each tree $\mathbb{G}_{j,n}^* = \{k \in \mathbb{G}_n : \delta_{(j,k)} = 1\}$ and $\mathbb{T}_{j,n}^* = \{k \in \mathbb{T}_n : \delta_{(j,k)} = 1\}$, and set

$$|\mathbb{G}_n^*| = \sum_{j=1}^m |\mathbb{G}_{j,n}^*| \quad \text{and} \quad |\mathbb{T}_n^*| = \sum_{j=1}^m |\mathbb{T}_{j,n}^*|,$$

the total number of observed cells in all m trees in generation n and up to generation n respectively. We next need to characterize the possible extinction of the GW processes, that is where $|\mathbb{T}_n^*|$ does not tend to infinity with n . For $1 \leq j \leq m$ and $n \geq 1$, we define the number of observed cells among the n -th generation of the j -th tree, distinguishing according to their type, by

$$Z_{j,n}^0 = \sum_{k \in \mathbb{G}_{n-1}} \delta_{(j,2k)} \quad \text{and} \quad Z_{j,n}^1 = \sum_{k \in \mathbb{G}_{n-1}} \delta_{(j,2k+1)},$$

and we set $\mathbf{Z}_{j,n} = (Z_{j,n}^0, Z_{j,n}^1)$. For all j , the process $(\mathbf{Z}_{j,n})$ thus defined is a two-type GW process, see [10]. We define the descendants matrix \mathbf{P} of the GW process by

$$\mathbf{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix},$$

where $p_{i0} = p^{(i)}(1, 0) + p^{(i)}(1, 1)$ and $p_{i1} = p^{(i)}(0, 1) + p^{(i)}(1, 1)$, for $i \in \{0, 1\}$. The quantity p_{il} is thus the expected number of descendants of type l of an individual of type i . It is well-known that when all the entries of the matrix \mathbf{P} are positive, \mathbf{P} has a positive strictly dominant eigenvalue, denoted π , which is also simple and admits a positive left eigenvector, see e.g. [10, Theorem 5.1]. In that case, we denote by $\mathbf{z} = (z^0, z^1)$ the left eigenvector of \mathbf{P} associated with the dominant eigenvalue π and satisfying $z^0 + z^1 = 1$. Let $\mathcal{E}_j = \bigcup_{n \geq 1} \{\mathbf{Z}_{j,n} = (0, 0)\}$ be the event corresponding to the case when there are no cells left to observe in the j -th tree. We will denote $\bar{\mathcal{E}}_j$ the complementary set of \mathcal{E}_j . We are interested in asymptotic results on the set where there is an infinity of $X_{(j,k)}$ to be observed that is on the union of the non-extinction sets $\bar{\mathcal{E}}_j$ denoted by

$$\bar{\mathcal{E}} = \bigcup_{j=1}^m \bar{\mathcal{E}}_j = \left\{ \lim_{n \rightarrow \infty} |\mathbb{T}_n^*| = \infty \right\}.$$

Note that we allow some trees to extinct, as long as there is at least one tree still growing.

A.2 Assumptions

Our inference is based on the m i.i.d. replicas of the observed BAR process, i.e. the available information is given by the sequence $(\delta_{(j,k)}, \delta_{(j,k)} X_{(j,k)})_{1 \leq j \leq m, k \geq 1}$. We first introduce the natural generation-wise filtrations of the BAR processes. For all $1 \leq j \leq m$, denote by $\mathbb{F}_j = (\mathcal{F}_{j,n})_{n \geq 1}$ the natural filtration associated with the j -th copy of the BAR process, which means that $\mathcal{F}_{j,n}$ is the σ -algebra generated by all individuals of the j -th tree up to the n -th generation, $\mathcal{F}_{j,n} = \sigma\{X_{(j,k)}, k \in \mathbb{T}_n\}$. For all $1 \leq j \leq m$, we also define the observation filtrations as $\mathcal{O}_{j,n} = \sigma\{\delta_{(j,k)}, k \in \mathbb{T}_n\}$, and the sigma fields $\mathcal{O}_j = \sigma\{\delta_{(j,k)}, k \geq 1\}$.

We make the following main assumptions on the BAR and GW processes.

(H.0) The parameters (a, b, c, d) satisfy the usual stability assumption $0 < \max\{|b|, |d|\} < 1$.

(H.1) For all $1 \leq j \leq m$, $n \geq 0$, $k \in \mathbb{G}_{n+1}$, $\mathbb{E}[\varepsilon_{(j,k)}^{16}] < \infty$ and $\mathbb{E}[X_{(j,1)}^{16}] < \infty$.

For all $1 \leq j \leq m$, $n \geq 0$ and $k \in \mathbb{G}_{n+1}$, one a.s. has

$$\begin{aligned} \mathbb{E}[\varepsilon_{(j,k)} | \mathcal{F}_{j,n}] = 0, \quad \mathbb{E}[\varepsilon_{(j,k)}^2 | \mathcal{F}_{j,n}] = \sigma^2, \quad \mathbb{E}[\varepsilon_{(j,k)}^3 | \mathcal{F}_{j,n}] = \lambda, \quad \mathbb{E}[\varepsilon_{(j,k)}^4 | \mathcal{F}_{j,n}] = \tau^4, \\ \mathbb{E}[\varepsilon_{(j,k)}^8 | \mathcal{F}_{j,n}] = \gamma^8, \quad \mathbb{E}[\varepsilon_{(j,k)}^{16} | \mathcal{F}_{j,n}] = \mu^{16}. \end{aligned}$$

For all $1 \leq j \leq m$, $n \geq 0$, $k \in \mathbb{G}_n$, one a.s. has

$$\begin{aligned} \mathbb{E}[\varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)} | \mathcal{F}_{j,n}] = \rho, \quad \mathbb{E}[\varepsilon_{(j,2k)}^2 \varepsilon_{(j,2k+1)}^2 | \mathcal{F}_{j,n}] = \nu^2, \quad \mathbb{E}[\varepsilon_{(j,2k)}^8 \varepsilon_{(j,2k+1)}^8 | \mathcal{F}_{j,n}] = \eta^8, \\ \mathbb{E}[\varepsilon_{(j,2k)}^2 \varepsilon_{(j,2k+1)} | \mathcal{F}_{j,n}] = \alpha, \quad \mathbb{E}[\varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)}^2 | \mathcal{F}_{j,n}] = \beta. \end{aligned}$$

(H.2) For all $1 \leq j \leq m$ and $n \geq 0$ the vectors $\{(\varepsilon_{(j,2k)}, \varepsilon_{(j,2k+1)}), k \in \mathbb{G}_n\}$ are conditionally independent given $\mathcal{F}_{j,n}$.

(H.3) The sequences $(\varepsilon_{(1,k)})_{k \geq 2}, (\varepsilon_{(2,k)})_{k \geq 2}, \dots, (\varepsilon_{(m,k)})_{k \geq 2}$ are independent. The random variables $(X_{(j,1)})_{1 \leq j \leq m}$ are independent and independent from the noise sequences.

(H.4) For all $1 \leq j \leq m$, the sequence $(\delta_{(j,k)})_{k \geq 1}$ is independent from the sequences $(X_{(j,k)})_{k \geq 1}$ and $(\varepsilon_{(j,k)})_{k \geq 2}$.

(H.5) The sequences $(\delta_{(1,k)})_{k \geq 2}, (\delta_{(2,k)})_{k \geq 2}, \dots, (\delta_{(m,k)})_{k \geq 2}$ are independent.

We also make the following super criticality assumption on the matrix \mathbf{P} .

(H.6) All entries of the matrix \mathbf{P} are positive: for all $(i, l) \in \{0, 1\}^2$, $p_{il} > 0$, and the dominant eigenvalue is greater than one: $\pi > 1$.

If $\pi > 1$, it is well known, see e.g. [10], that the extinction probability of the GW processes is less than one: for all $1 \leq j \leq m$, $\mathbb{P}(\mathcal{E}_j) = p < 1$. Under assumptions (H.5-6), one thus clearly has $\mathbb{P}(\bar{\mathcal{E}}) = 1 - p^m > 0$.

A.3 Additional estimators

From the estimators of the reproductions probabilities of the GW process, one can easily construct an estimator of the spectral radius π of the descendants matrix \mathbf{P} of the GW process. Indeed, \mathbf{P} is a 2×2 matrix so that its spectral radius can be computed explicitly as a function of its coefficients, namely

$$\pi = \frac{1}{2} \left(\text{tr}(\mathbf{P}) + (\text{tr}(\mathbf{P})^2 - 4 \det(\mathbf{P}))^{1/2} \right).$$

Replacing the coefficients of \mathbf{P} by their empirical estimators, one obtains

$$\hat{\pi}_n = \frac{1}{2} (\hat{T}_n + (\hat{T}_n^2 - 4\hat{D}_n)^{1/2}).$$

where

$$\begin{aligned} \hat{T}_n &= \hat{p}_n^{(0)}(1, 0) + \hat{p}_n^{(0)}(1, 1) + \hat{p}_n^{(1)}(0, 1) + \hat{p}_n^{(1)}(1, 1), \\ \hat{D}_n &= (\hat{p}_n^{(0)}(1, 0) + \hat{p}_n^{(0)}(1, 1))(\hat{p}_n^{(1)}(0, 1) + \hat{p}_n^{(1)}(1, 1)) - (\hat{p}_n^{(0)}(0, 1) + \hat{p}_n^{(0)}(1, 1))(\hat{p}_n^{(1)}(1, 0) + \hat{p}_n^{(1)}(1, 1)) \end{aligned}$$

are the empirical estimator of the trace $\text{tr}(\mathbf{P})$ and the determinant $\det(\mathbf{P})$ respectively. Finally, to compute confidence intervals for σ^2 and ρ , we need an estimation of higher moments. We use again empirical estimators

$$\hat{\tau}_n^4 = \frac{1}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} (\hat{\varepsilon}_{(j,2k)}^4 + \hat{\varepsilon}_{(j,2k+1)}^4), \quad \hat{\nu}_n^2 = \frac{1}{|\mathbb{T}_{n-1}^{*01}|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k)}^2 \hat{\varepsilon}_{(j,2k+1)}^2.$$

B Convergence of estimators for the GW process

We now prove the convergence of the estimators for the GW process, that is the first parts of Theorems 3.1 and 3.2, together with additional technical results.

B.1 Preliminary results: from single trees to multiple trees

Our first objective is to show that we can adapt the results in [3] to the multiple tree framework. To this aim, we first need to recall Lemma A.3 of [1].

Lemma B.1 *Let (\mathbf{A}_n) be a sequence of real-valued matrices such that*

$$\sum_{n=0}^{\infty} \|\mathbf{A}_n\| < \infty \quad \text{and} \quad \lim_{n \rightarrow \infty} \sum_{k=0}^n \mathbf{A}_k = \mathbf{A}.$$

In addition, let (\mathbf{X}_n) be a sequence of real-valued vectors which converges to a limiting value \mathbf{X} . Then, one has

$$\lim_{n \rightarrow \infty} \sum_{\ell=0}^n \mathbf{A}_{n-\ell} \mathbf{X}_\ell = \mathbf{A} \mathbf{X}.$$

The next result is an adaptation of Lemma A.2 in [1] to the GW tree framework. It gives a correspondence between sums on one generation and sums on the whole tree.

Lemma B.2 *Let (x_n) be a sequence of real numbers and $\pi > 1$. One has*

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k = x \iff \lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{G}_n} x_k = \frac{\pi - 1}{\pi} x.$$

Proof: Suppose that $\pi^{-n} \sum_{k \in \mathbb{T}_n} x_k$ converges to x . Then one has

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{G}_n} x_k = \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k - \frac{1}{\pi} \frac{1}{\pi^{n-1}} \sum_{k \in \mathbb{T}_{n-1}} x_k \xrightarrow{n \rightarrow \infty} x - \frac{1}{\pi} x = \frac{\pi - 1}{\pi} x.$$

Conversely, if $\pi^{-n} \sum_{k \in \mathbb{G}_n} x_k$ converges to y , as $\mathbb{T}_n = \cup_{\ell=0}^n \mathbb{G}_\ell$, one has

$$\frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_k = \sum_{\ell=0}^n \frac{1}{\pi^{n-\ell}} \frac{1}{\pi^\ell} \sum_{k \in \mathbb{G}_\ell} x_k \xrightarrow{n \rightarrow \infty} \frac{\pi}{\pi-1} y,$$

using Lemma B.1 with $\mathbf{A}_n = \pi^{-n}$ and $\mathbf{X}_n = \pi^{-n} \sum_{k \in \mathbb{G}_n} x_k$. \square

We now adapt Lemma 2.1 of [3] to our multiple tree framework.

Lemma B.3 *Under assumption (H.5-6), there exist a nonnegative random variable W such that for all sequences $(x_{(1,n)}), \dots, (x_{(m,n)})$ of real numbers one has a.s.*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = x \mathbb{1}_{\bar{\mathcal{E}}} \iff \lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = x \frac{\pi}{\pi-1} W.$$

Proof: We use a well known property of super-critical GW processes, see e.g. [10]: for all j , there exists a non negative random variable W_j such that

$$\lim_{n \rightarrow \infty} \frac{|\mathbb{T}_{j,n}^*|}{\pi^n} = \frac{\pi}{\pi-1} W_j \quad \text{a.s.} \quad (4)$$

and in addition $\{W_j > 0\} = \bar{\mathcal{E}}_j = \lim\{|\mathbb{G}_{j,n}^*| > 0\}$. Therefore, one has

$$\lim_{n \rightarrow \infty} \sum_{j=1}^m \frac{|\mathbb{T}_{j,n}^*|}{\pi^n} = \lim_{n \rightarrow \infty} \frac{|\mathbb{T}_n^*|}{\pi^n} = \frac{\pi}{\pi-1} \sum_{j=1}^m W_j \quad \text{a.s.}$$

The result is obtained by setting $W = \sum_{j=1}^m W_j$ and noticing that $\bar{\mathcal{E}} = \cup_{j=1}^m \bar{\mathcal{E}}_j = \{\sum_{j=1}^m W_j > 0\} = \lim\{|\mathbb{G}_n^*| > 0\}$. \square

Finally, the main result of this section is new and explains how convergence results on multiple trees can be obtained from convergence results on a single tree. This will allow us to directly use results from [3] in all the sequel.

Lemma B.4 *Let $(x_{(1,n)}), \dots, (x_{(m,n)})$ be m sequences of real numbers such that for all $1 \leq j \leq m$ one has the a.s. limit*

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_{j,n}^*| > 0\}}}{|\mathbb{T}_{j,n}^*|} \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \mathbb{1}_{\bar{\mathcal{E}}_j}, \quad (5)$$

then under assumptions (H.5-6) one also has

$$\lim_{n \rightarrow \infty} \frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \mathbb{1}_{\bar{\mathcal{E}}} \quad \text{a.s.}$$

Proof: Equations (5) and (4) yield, for all j ,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \frac{\pi-1}{\pi} W_j.$$

Summing over j , one obtains

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} x_{(j,k)} = \ell \frac{\pi-1}{\pi} \sum_{j=1}^m W_j = \ell \frac{\pi-1}{\pi} W.$$

Finally, we use Lemma B.3 to conclude. \square

B.2 Strong consistency for the estimators of the GW process

To prove the convergence of the $\widehat{p}_n^{(i)}(l_0, l_1)$ we first need to derive a convergence result for a sum of independent GW processes.

Lemma B.5 *Suppose that assumptions (H.5-6) are satisfied. Then for $i \in \{0, 1\}$ one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} = z^i \mathbb{1}_{\overline{\mathcal{E}}} \quad \text{a.s.}$$

Proof Remarking that $\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} = \sum_{j=1}^m \sum_{l=1}^n Z_{j,l}^i$, the lemma is a direct consequence of Lemma B.4 and the well-known property of super-critical GW processes $\mathbb{1}_{\{|\mathbb{G}_{j,n}^*| > 0\}} |\mathbb{T}_{j,n}^*|^{-1} \sum_{l=1}^n \mathbf{Z}_{j,l} \rightarrow \mathbf{z} \mathbb{1}_{\overline{\mathcal{E}}}$, for all $0 \leq j \leq m$. \square

Proof of Theorem 3.1, first part We give the details of the convergence of $\widehat{p}_n^{(1)}(1, 1)$ to $p^{(1)}(1, 1)$, the other convergences are derived similarly. The proof relies on the convergence of square integrable scalar martingales. Set

$$M_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)} (\delta_{(j, 4k+2)} \delta_{(j, 4k+3)} - p^{(1)}(1, 1)).$$

We are going to prove that (M_n) is a martingale for a well chosen filtration. Recall that $\mathcal{O}_{j,n} = \sigma\{\delta_{(j,k)}, k \in \mathbb{T}_n\}$, and set $\mathcal{O}_n = \vee_{j=1}^m \mathcal{O}_{j,n}$. Then (M_n) is clearly a square integrable real (\mathcal{O}_n) -martingale. Using the independence assumption (H.5), its increasing process is

$$\langle M \rangle_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)} p^{(1)}(1, 1) (1 - p^{(1)}(1, 1)) = p^{(1)}(1, 1) (1 - p^{(1)}(1, 1)) \sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1.$$

Hence, Lemma B.5 implies that $|\mathbb{T}_{n-1}^*|^{-1} \langle M \rangle_n$ converges almost surely on the non extinction set $\overline{\mathcal{E}}$. The law of large numbers for scalar martingales thus yields that $|\mathbb{T}_{n-1}^*|^{-1} M_n$ tends to 0 as n tends to infinity on $\overline{\mathcal{E}}$. Finally, notice that

$$\widehat{p}_n^{(1)}(1, 1) - p^{(1)}(1, 1) = \frac{M_n}{\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j, 2k+1)}} = \frac{M_n}{\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1},$$

so that Lemma B.5 again implies the almost sure convergence of $\widehat{p}_n^{(1)}(1, 1)$ to $p^{(1)}(1, 1)$ on the non extinction set $\overline{\mathcal{E}}$. \square

As a direct consequence, one obtains the a.s. convergence of $\widehat{\pi}_n$ to π on $\overline{\mathcal{E}}$.

B.3 Asymptotic normality for the estimators of the GW process

As $\mathbb{P}(\overline{\mathcal{E}}) \neq 0$, we can define a new probability $\mathbb{P}_{\overline{\mathcal{E}}}$ by $\mathbb{P}_{\overline{\mathcal{E}}}(A) = \mathbb{P}(A \cap \overline{\mathcal{E}}) / \mathbb{P}(\overline{\mathcal{E}})$ for all event A . In all the sequel of this section, we will work on the space $\overline{\mathcal{E}}$ under the probability $\mathbb{P}_{\overline{\mathcal{E}}}$ and we denote by $\mathbb{E}_{\overline{\mathcal{E}}}$ the corresponding expectation. We can now turn to the proof of the asymptotic normality of \widehat{p}_n . The proof also relies on martingale theory. As the normalizing term in our central limit theorem is random, we use the central limit theorem for martingales given in Theorem 2.1.9 of [6] that we first recall as Theorem B.6 for self-completeness.

Theorem B.6 *Suppose that (Ω, \mathcal{A}, P) is a probability space and that for each n we have a filtration $\mathbb{F}_n = (\mathcal{F}_k^{(n)})$, a stopping time ν_n relative to \mathbb{F}_n and a real, square-integrable vector martingale $M^{(n)} = (M_k^{(n)})_{k \geq 0}$ which is adapted to \mathbb{F}_n and has hook denoted by $\langle M \rangle^{(n)}$. We make the following two assumptions.*

A.1 *For a deterministic symmetric positive semi-definite matrix Γ*

$$\langle M \rangle_{\nu_n}^{(n)} \xrightarrow{P} \Gamma.$$

A.2 *Lindeberg's condition holds; in other words, for all $\epsilon > 0$,*

$$\sum_{k=1}^{\nu_n} \mathbb{E} \left(\|M_k^{(n)} - M_{k-1}^{(n)}\|^2 \mathbb{1}_{\{\|M_k^{(n)} - M_{k-1}^{(n)}\| > \epsilon\}} \mid \mathcal{F}_{k-1}^{(n)} \right) \xrightarrow{P} 0.$$

Then:

$$M_{\nu_n}^{(n)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma).$$

Proof of Theorem 3.2, first part First, set

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}^0/z^0 & 0 \\ 0 & \mathbf{V}^1/z^1 \end{pmatrix} \quad (6)$$

where for all i in $\{0, 1\}$, $\mathbf{V}^i = \mathbf{W}^i - \mathbf{p}^{(i)}(\mathbf{p}^{(i)})^t$, \mathbf{W}^i is a 4×4 matrix with the entries of $\mathbf{p}^{(i)}$ on the diagonal and 0 elsewhere. We are going to prove that \mathbf{V} is the asymptotic variance of $\widehat{\mathbf{p}}_n - \mathbf{p}$ suitably normalized. We use Theorem B.6. We first need to define a suitable filtration. Here, we use the first cousins filtration defined as follows. Let

$$\mathcal{H}_{j,p} = \sigma\{\delta_{(j,1)}, \dots, \delta_{(j,3)}, (\delta_{(j,4k)}, \dots, \delta_{(j,4k+3)}), 1 \leq k \leq p\}$$

be the σ -field generated by all the 4-tuples of observed cousin cells up the granddaughters of cell (j, p) in the j -th tree and $\mathcal{H}_p = \bigvee_{j=1}^m \mathcal{H}_{j,p}$. Hence, the 4-tuple $(\delta_{(j,4k)}, \dots, \delta_{(j,4k+3)})$ is \mathcal{H}_k -measurable for all j . By definition of the reproduction probabilities $p^{(i)}(l_0, l_1)$, the processes

$$(\delta_{(j,2k+i)}(\phi_{l_0}(\delta_{(j,2(2k+i))})\phi_{l_1}(\delta_{(j,2(2k+i)+1})) - p^{(i)}(l_0, l_1))_{k \geq 1}$$

are (\mathcal{H}_k) -martingale difference sequences. We thus introduce a sequence of (\mathcal{H}_k) -martingales $(\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$ defined for all $n \geq 1$ and $p \geq 1$ by

$$\mathbf{M}_p^{(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m \mathbf{D}_{(j,k)},$$

with $\mathbf{D}_{(j,k)} = ((\mathbf{D}_{(j,k)}^0)^t, (\mathbf{D}_{(j,k)}^1)^t)^t$ and

$$\mathbf{D}_{(j,k)}^i = \delta_{(j,2k+i)} \begin{pmatrix} (1 - \delta_{(j,2(2k+i))})(1 - \delta_{(j,2(2k+i)+1)}) - p^{(i)}(0, 0) \\ (1 - \delta_{(j,2(2k+i))})\delta_{(j,2(2k+i)+1)} - p^{(i)}(0, 1) \\ \delta_{(j,2(2k+i))}(1 - \delta_{(j,2(2k+i)+1)}) - p^{(i)}(1, 0) \\ \delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)} - p^{(i)}(1, 1) \end{pmatrix}.$$

We also introduce the sequence of stopping times $\nu_n = |\mathbb{T}_{n-2}| = 2^{n-1} - 1$. One has

$$\mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_{(j,k)} \mathbf{D}_{(j,k)}^t | \mathcal{H}_{k-1}] = \begin{pmatrix} \delta_{(j,2k)} \mathbf{V}^0 & 0 \\ 0 & \delta_{(j,2k+1)} \mathbf{V}^1 \end{pmatrix}.$$

Therefore the one has $\langle \mathbf{M}^{(n)} \rangle_{\nu_n} = |\mathbb{T}_{n-1}^*|^{-1} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \begin{pmatrix} Z_{j,\ell}^0 \mathbf{V}^0 & 0 \\ 0 & Z_{j,\ell}^1 \mathbf{V}^1 \end{pmatrix}$, so that its $\mathbb{P}_{\bar{\mathcal{E}}}$ almost sure limit is

$$\Gamma' = \begin{pmatrix} z^0 \mathbf{V}^0 & 0 \\ 0 & z^1 \mathbf{V}^1 \end{pmatrix},$$

thanks to Lemma B.5. Therefore, assumption A.1 of Theorem B.6 holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. The Lindeberg condition A.2 is obviously satisfied as we deal with finite support distributions. We then conclude that under $\mathbb{P}_{\bar{\mathcal{E}}}$ one has

$$|\mathbb{T}_{n-1}^*|^{-1/2} \mathbf{M}_{\nu_n}^{(n)} = |\mathbb{T}_{n-1}^*|^{-1/2} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}^*} \mathbf{D}_{(j,k)} \xrightarrow{\mathcal{L}} \mathcal{N}(0, \Gamma').$$

Using the relation

$$\widehat{\mathbf{p}}_n - \mathbf{p} = \begin{pmatrix} (\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^0) \mathbf{I}_4 & 0 \\ 0 & (\sum_{j=1}^m \sum_{\ell=0}^{n-1} Z_{j,\ell}^1) \mathbf{I}_4 \end{pmatrix}^{-1} \mathbf{M}_{\nu_n}^{(n)},$$

Lemma B.5 and Slutsky's Lemma give the first part of Theorem 3.2. \square

B.4 Interval estimation and tests for the GW process

From the central limit theorem 3.2 one can easily build asymptotic confidence intervals for our estimators. In our context, Y_n and Y'_n being two random variables, we will say that $[Y_n; Y'_n]$ is an asymptotic confidence interval with confidence level $1 - \epsilon$ for the parameter Y if $\mathbb{P}_{\bar{\mathcal{E}}}(Y_n \leq Y \leq Y'_n) \xrightarrow{n \rightarrow \infty} (1 - \epsilon)$. For any $0 \leq \epsilon \leq 1$, let $q_{1-\epsilon/2}$ be the $1 - \epsilon/2$ quantile of the standard normal law.

For all $n \geq 2$, define the 8×8 matrix

$$\widehat{\mathbf{V}}_n = \begin{pmatrix} \widehat{\mathbf{V}}_n^0 (\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k)})^{-1} & 0 \\ 0 & \widehat{\mathbf{V}}_n^1 (\sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-2}} \delta_{(j,2k+1)})^{-1} \end{pmatrix}$$

where for all i in $\{0, 1\}$, $\widehat{\mathbf{V}}_n^i = \widehat{\mathbf{W}}_n^i - \widehat{\mathbf{p}}_n^{(i)} (\widehat{\mathbf{p}}_n^{(i)})^t$, $\widehat{\mathbf{W}}_n^i$ is a 4×4 matrix with the entries of $\widehat{\mathbf{p}}_n^{(i)}$ on the diagonal and 0 elsewhere. Thus, $|\mathbb{T}_{n-1}^*| \widehat{\mathbf{V}}_n$ is an empirical estimator of the covariance matrix \mathbf{V} .

Theorem B.7 *Under assumptions (H.5-6), for i, l_0, l_1 in $\{0, 1\}$ and for any $0 < \epsilon < 1$, the random interval defined by*

$$[\widehat{p}_n^{(i)}(l_0, l_1) - q_{1-\epsilon/2} (\widehat{\mathbf{V}}_n^{1/2})_{\ell, \ell} ; \widehat{p}_n^{(i)}(l_0, l_1) + q_{1-\epsilon/2} (\widehat{\mathbf{V}}_n^{1/2})_{\ell, \ell}]$$

is an asymptotic confidence interval with level $1 - \epsilon$ for $p^{(i)}(l_0, l_1)$; where (ℓ, ℓ) is the coordinate of \mathbf{V}_n corresponding to $p^{(i)}(l_0, l_1)$, namely $\ell = 4i + 1 + 2l_0 + l_1$.

Proof This is a straightforward consequence of the central limit Theorem 3.2 together with Stulsky's lemma as $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^*| \widehat{\mathbf{V}}_n = \mathbf{V}$ $\mathbb{P}_{\bar{\mathcal{E}}}$ a.s. thanks to Lemma B.5 and Theorem 3.1. \square

Set $\widehat{G}_n = \widehat{\mathbf{F}}_n^t \widehat{\mathbf{V}}_n \widehat{\mathbf{F}}_n$, where $\widehat{\mathbf{F}}_n$ is the 8×1 vector defined by

$$\widehat{\mathbf{F}}_n = \frac{1}{2} (0, 0, 1, 1, 0, 1, 0, 1)^t + \frac{1}{2} (\widehat{T}_n^2 - 4\widehat{D}_n)^{-1/2} \widehat{\mathbf{H}}_n$$

and

$$\widehat{\mathbf{H}}_n = \begin{pmatrix} 0 \\ 2\widehat{p}_n^{(1)}(1, 0) + 2\widehat{p}_n^{(1)}(1, 1) \\ \widehat{p}_n^{(0)}(1, 0) + \widehat{p}_n^{(0)}(1, 1) - \widehat{p}_n^{(1)}(0, 1) - \widehat{p}_n^{(1)}(1, 1) \\ \widehat{p}_n^{(0)}(1, 0) + \widehat{p}_n^{(0)}(1, 1) - \widehat{p}_n^{(1)}(0, 1) + 2\widehat{p}_n^{(1)}(1, 0) + \widehat{p}_n^{(1)}(1, 1) \\ 0 \\ -\widehat{p}_n^{(0)}(1, 0) - \widehat{p}_n^{(0)}(1, 1) + \widehat{p}_n^{(1)}(0, 1) + \widehat{p}_n^{(1)}(1, 1) \\ 2\widehat{p}_n^{(0)}(0, 1) + 2\widehat{p}_n^{(0)}(1, 1) \\ 2\widehat{p}_n^{(0)}(0, 1) - \widehat{p}_n^{(0)}(1, 0) + \widehat{p}_n^{(0)}(1, 1) + \widehat{p}_n^{(1)}(0, 1) + \widehat{p}_n^{(1)}(1, 1) \end{pmatrix}.$$

Theorem B.8 *Under assumptions (H.5-6), for any $0 < \epsilon < 1$ one has that*

$$[\widehat{\pi}_n - q_{1-\epsilon/2} \widehat{G}_n^{1/2} ; \widehat{\pi}_n + q_{1-\epsilon/2} \widehat{G}_n^{1/2}]$$

is an asymptotic confidence interval with level $1 - \epsilon$ for π .

Proof This is again a straightforward consequence of the central limit Theorem 3.2 together with Stulsky's lemma as $\widehat{\mathbf{F}}_n$ is the gradient of the function that maps the vector $\widehat{\mathbf{p}}$ onto the estimator $\widehat{\pi}_n$. \square

We propose two symmetry tests for the GW process. The first one compares the average number of offspring m_0 of a cell of type 0: $m_0 = p^{(0)}(1, 0) + p^{(0)}(0, 1) + 2p^{(0)}(1, 1)$ to that of a cell of type 1: $m_1 = p^{(1)}(1, 0) + p^{(1)}(0, 1) + 2p^{(1)}(1, 1)$. Denote by \widehat{m}_n^0 and \widehat{m}_n^1 their empirical estimators. Set

- \mathbf{H}_0^m : $m_0 = m_1$ the symmetry hypothesis,

- \mathbf{H}_1^m : $m_0 \neq m_1$ the alternative hypothesis.

Let Y_n^m be the test statistic defined by

$$Y_n^m = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^m)^{-1/2} (\widehat{m}_n^0 - \widehat{m}_n^1),$$

where $\widehat{\Delta}_n^m = |\mathbb{T}_{n-1}^*| \mathbf{d}\mathbf{g}_m^t \widehat{\mathbf{V}}_n \mathbf{d}\mathbf{g}_m$ and $\mathbf{d}\mathbf{g}_m = (0, 1, 1, 2, 0 - 1, -1, -2)^t$. This test statistic has the following asymptotic properties.

Theorem B.9 *Under assumptions (H.5-6) and the null hypothesis \mathbf{H}_0^m , one has*

$$(Y_n^m)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^m , almost surely on $\bar{\mathcal{E}}$ one has

$$\lim_{n \rightarrow \infty} (Y_n^m)^2 = +\infty.$$

Proof Let g_m be the function defined from \mathbb{R}^8 onto \mathbb{R} by $g_m(x_1, \dots, x_8) = x_2 + x_3 + 2x_4 - x_6 - x_7 - 2x_8$ so that $\mathbf{d}\mathbf{g}_m$ is the gradient of g_m . Thus, the central limit Theorem 3.2 yields

$$\sqrt{|\mathbb{T}_{n-1}^*|} (g_m(\widehat{\mathbf{p}}_n) - g_m(\mathbf{p})) \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{d}\mathbf{g}_m^t \mathbf{V} \mathbf{d}\mathbf{g}_m) = \mathcal{N}(0, \Delta^m)$$

on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. Under the null hypothesis \mathbf{H}_0^m , $g_m(\mathbf{p}) = 0$, so that one has

$$|\mathbb{T}_{n-1}^*| (\Delta^m)^{-1} g_m(\widehat{\mathbf{p}}_n)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. Lemma B.5 and Theorem 3.1 give the almost sure convergence of $\widehat{\Delta}_n^m$ to Δ^m . Hence Slutsky's Lemma yields the expected result. Under the alternative hypothesis \mathbf{H}_1^m , one has

$$Y_n^m = (\widehat{\Delta}_n^m)^{-1/2} (\sqrt{|\mathbb{T}_{n-1}^*|} (g_m(\widehat{\mathbf{p}}_n) - g_m(\mathbf{p})) + \sqrt{|\mathbb{T}_{n-1}^*|} g_m(\mathbf{p})).$$

The first term converges to a centered normal law and the second term tends to infinity as $|\mathbb{T}_{n-1}^*|$ tends to infinity a.s. on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$. \square

Our next test compares the reproduction probability vectors of mother cells of type 0 and 1.

- \mathbf{H}_0^p : $\mathbf{p}^{(0)} = \mathbf{p}^{(1)}$ the symmetry hypothesis,
- \mathbf{H}_1^p : $\mathbf{p}^{(0)} \neq \mathbf{p}^{(1)}$ the alternative hypothesis.

Let $(\mathbf{Y}_n^p)^t \mathbf{Y}_n^p$ be the test statistic defined by

$$\mathbf{Y}_n^p = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^p)^{-1/2} (\widehat{\mathbf{p}}^{(0)} - \widehat{\mathbf{p}}^{(1)}),$$

where $\widehat{\Delta}_n^p = |\mathbb{T}_{n-1}^*| \mathbf{d}\mathbf{g}_p^t \widehat{\mathbf{V}}_n \mathbf{d}\mathbf{g}_p$ and $\mathbf{d}\mathbf{g}_p = \begin{pmatrix} \mathbf{I}_4 \\ -\mathbf{I}_4 \end{pmatrix}$. This test statistic has the following asymptotic properties.

Theorem B.10 *Under assumptions (H.5-6) and the null hypothesis \mathbf{H}_0^p , one has*

$$(\mathbf{Y}_n^p)^t \mathbf{Y}_n^p \xrightarrow{\mathcal{L}} \chi^2(4)$$

on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^p , almost surely on $\bar{\mathcal{E}}$ one has

$$\lim_{n \rightarrow \infty} \|\mathbf{Y}_n^p\|^2 = +\infty.$$

Proof We mimic the proof of Theorem B.9 with g_p the function defined from \mathbb{R}^8 onto \mathbb{R}^4 by $g_p(x_1, \dots, x_8) = (x_1 - x_5, x_2 - x_6, x_3 - x_7, x_4 - x_8)^t$, so that $\mathbf{d}\mathbf{g}_p$ is the gradient of g_p . \square

C Convergence of estimators for the BAR process

We now prove the convergence of the estimators for the BAR process, that is the parts of Theorems 3.1 and 3.2 concerning $\widehat{\theta}_n$, $\widehat{\sigma}_n^2$ and $\widehat{\rho}_n$, together with additional technical results, especially the convergence of higher moment estimators required to estimate the asymptotic variances.

C.1 Preliminary results: laws of large numbers

In this section, we want to study the asymptotic behavior of various sums of observed data. Most of the results are directly taken from [3]. All external references in this section refer to that paper that will not be cited each time. However, we need additional results concerning higher moments of the BAR process in order to obtain the consistency of $\widehat{\tau}_n^4$ and $\widehat{\nu}_n^2$, as there is no such result in [3]. We also give all the explicit formulas so that the interested reader can actually compute the various asymptotic variances.

Again, our work relies on the strong law of large numbers for square integrable martingales. To ensure that the increasing processes of our martingales are at most $\mathcal{O}(\pi^n)$ we first need the following lemma.

Lemma C.1 *Under assumptions (H.0-6), for all $i \in \{0, 1\}$ one has*

$$\sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^8 = \mathcal{O}(\pi^n) \quad a.s.$$

Proof The proof follows the same lines as that of Lemma 6.1. The constants before the terms A_n^i , B_n^i and C_n^i therein are replaced respectively by $(4/(1-\beta))^7$, $\alpha^8(4/(1-\beta))^7$ and 2^8 ; in the term A_n^i , ε^2 is replaced by ε^8 ; in the term C_n^i , β^{2r_k} is replaced by β^{8r_k} ; the term B_n^i is unchanged. In the expression of $\mathbb{E}[(Y_{\ell,p}^i)^2]$, one just needs to replace τ^4 by μ^{16} , σ^4 by γ^{16} and $\nu^2\tau^4$ by η^8 . Note that the various moments of the noise sequence are defined in assumption (H.1). The rest of the proof is unchanged. \square

We also state some laws of large numbers for the noise processes.

Lemma C.2 *Under assumptions (H.0-6), for all $i \in \{0, 1\}$ and for all integers $0 \leq q \leq 4$, one has*

$$\frac{1}{\pi^n} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+i)} \varepsilon_{(j, 2k+i)}^q = \frac{\pi}{\pi-1} W z^i \mathbb{E}[\varepsilon_{(1, 2+i)}^q] \quad a.s.$$

Proof This is also a direct consequence of [3] thanks to Lemmas B.3 and B.4. Lemma 5.3 provides the result for $q = 0$, Lemma 5.5 for $q = 1$, Corollary 5.6 for $q = 2$ and Lemma 5.7 for $q = 4$. The result for $q = 3$ is obtained similarly. \square

In view of these new stronger results, we can now state our first laws of large numbers for the observed BAR process. For $i \in \{0, 1\}$ and all integers $1 \leq q \leq 4$ let us now define

$$H_n^i(q) = \sum_{j=1}^m H_{j,n}^i(q) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^q, \quad H_n^{01}(q) = \sum_{j=1}^m H_{j,n}^{01}(q) = \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k)} \delta_{(j, 2k+1)} X_{(j,k)}^q,$$

and $\mathbf{H}_n(q) = (H_n^0(q), H_n^1(q))^t$.

Lemma C.3 *Under assumptions (H.0-6) and for all integers $1 \leq q \leq 4$, one has the following a.s. limits on the non extinction set $\bar{\mathcal{E}}$*

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \mathbf{H}_n(q) &= \mathbf{h}(q) = (\mathbf{I}_2 - \widetilde{\mathbf{P}}_q)^{-1} \mathbf{P}^t \widetilde{\mathbf{h}}(q), \\ \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} H_n^{01}(q) &= h^{01}(q) = p^{(0)}(1, 1) \left(\widetilde{h}^0(q) + b^q \frac{h^0(q)}{\pi} \right) + p^{(1)}(1, 1) \left(\widetilde{h}^1(q) + d^q \frac{h^1(q)}{\pi} \right), \end{aligned}$$

where

$$\tilde{\mathbf{P}}_q = \pi^{-1} \mathbf{P}^t \begin{pmatrix} b^q & 0 \\ 0 & d^q \end{pmatrix}, \quad \mathbf{h}(q) = \begin{pmatrix} h^0(q) \\ h^1(q) \end{pmatrix}, \quad \tilde{\mathbf{h}}(q) = \begin{pmatrix} \tilde{h}^0(q) \\ \tilde{h}^1(q) \end{pmatrix},$$

and

$$\begin{aligned} \tilde{h}^0(1) &= az^0, & \tilde{h}^1(1) &= cz^1, \\ \tilde{h}^0(2) &= (a^2 + \sigma^2)z^0 + 2abh^0(1)\pi^{-1}, & \tilde{h}^1(2) &= (c^2 + \sigma^2)z^1 + 2cdh^1(1)\pi^{-1}, \\ \tilde{h}^0(3) &= (a^3 + 3a\sigma^2 + \lambda)z^0 + 3b(a^2 + \sigma^2)h^0(1)\pi^{-1} + 3ab^2h^0(2)\pi^{-1}, \\ \tilde{h}^1(3) &= (c^3 + 3c\sigma^2 + \lambda)z^1 + 3d(c^2 + \sigma^2)h^1(1)\pi^{-1} + 3cd^2h^1(2)\pi^{-1}, \\ \tilde{h}^0(4) &= (a^4 + 6a^2\sigma^2 + 4a\lambda + \tau^4)z^0 + 4b(a^3 + 3a\sigma^2 + \lambda)h^0(1)\pi^{-1} + 6b^2(a^2 + \sigma^2)h^0(2)\pi^{-1} + 4ab^3h^0(3)\pi^{-1}, \\ \tilde{h}^1(4) &= (c^4 + 6c^2\sigma^2 + 4c\lambda + \tau^4)z^1 + 4d(c^3 + 3c\sigma^2 + \lambda)h^1(1)\pi^{-1} + 6d^2(c^2 + \sigma^2)h^1(2)\pi^{-1} + 4cd^3h^1(3)\pi^{-1}. \end{aligned}$$

Proof The results for $q = 1$ and $q = 2$ come from Propositions 6.3, 6.5 and 6.6 together with Lemma B.4. The proofs for $q \geq 3$ follow the same lines, using Lemma C.2 when required and Lemma C.1 to bound the increasing processes of the various martingales at stake. \square

To prove the consistency of our estimators, we also need some additional families of laws of large numbers.

Lemma C.4 *Under assumptions (H.0-6), for $i \in \{0, 1\}$ and for all integers $1 \leq p + q \leq 4$, one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p \varepsilon_{(j, 2k+i)}^q = \mathbb{E}[\varepsilon_{2+i}^q] h^i(p) \mathbb{1}_{\bar{\mathcal{E}}}.$$

Proof The proof is similar to that of Theorem 3.1. For all $1 \leq j \leq m$, one has

$$\sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p \varepsilon_{(j, 2k+i)}^q = \sum_{\ell=0}^n \sum_{k \in \mathbb{G}_\ell} \delta_{(j, 2k+i)} X_{(j,k)}^p (\varepsilon_{(j, 2k+i)}^q - \mathbb{E}[\varepsilon_{(j, 2k+i)}^q | \mathcal{F}_{j,\ell}^{\mathcal{O}}]) + \mathbb{E}[\varepsilon_{2+i}^q] \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p,$$

as the conditional moment of ε_{2k+i} are constants by assumption (H.1). The first term is a square integrable $(\mathcal{F}_{j,n}^{\mathcal{O}})$ -martingale and its increasing process is $\mathcal{O}(\pi^n)$ thanks to Lemma C.1, thus the first term is $o(\pi^n)$. The limit of the second term is given by Lemma C.3. \square

Lemma C.5 *Under assumptions (H.0-6), for $i \in \{0, 1\}$ and for all integers $1 \leq q \leq 4$, one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j, 2k+i)}^q = (\pi \tilde{h}^i(q) + b^q h^i(q)) \mathbb{1}_{\bar{\mathcal{E}}}.$$

Proof The proof is obtained by replacing $X_{(j, 2k+i)}$ by $a + bX_k + \varepsilon_{2k}$ if $i = 0$ or $c + dX_k + \varepsilon_{2k+1}$ if $i = 1$. One then develops the exponent and uses Lemmas B.5, C.2, C.3 and C.4 to conclude. \square

Lemma C.6 *Under assumptions (H.0-6), for $i \in \{0, 1\}$ and for all integers $1 \leq p + q \leq 4$, one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_n} \delta_{(j, 2k+i)} X_{(j,k)}^p X_{(j, 2k+i)}^q = h^i(p, q) \mathbb{1}_{\bar{\mathcal{E}}},$$

with

$$\begin{aligned} h^0(p, 1) &= ah^0(p) + bh^0(p+1), & h^1(p, 1) &= ch^1(p) + dh^1(p+1), \\ h^0(p, 2) &= (a^2 + \sigma^2)h^0(p) + 2abh^0(p+1) + b^2h^0(p+2), \\ h^1(p, 2) &= (c^2 + \sigma^2)h^1(p) + 2cdh^1(p+1) + d^2h^1(p+2), \\ h^0(p, 3) &= (a^3 + 3a\sigma^2 + \lambda)h^0(p) + 3b(a^2 + \sigma^2)h^0(p+1) + 3ab^2h^0(p+2) + b^3h^0(p+3), \\ h^1(p, 3) &= (c^3 + 3c\sigma^2 + \lambda)h^1(p) + 3d(c^2 + \sigma^2)h^1(p+1) + 3cd^2h^1(p+2) + d^3h^1(p+3), \end{aligned}$$

where we used the convention $h^i(0) = z^i \pi$.

Proof As above, the proof is obtained by replacing $X_{(j,2k+i)}$ and developing the exponents. Then one uses Lemmas B.5, C.2, C.3 and C.4 to compute the limits. \square

Lemma C.7 *Under assumptions (H.5-6), one has the following a.s. limit*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1}\sum_{j=1}^m\sum_{k\in\mathbb{T}_n}\delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)}=p^{(i)}(1,1)z^i\mathbb{1}_{\bar{\mathcal{E}}}.$$

Proof First note that $\delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)}=\delta_{(j,2k+i)}\delta_{(j,2(2k+i))}\delta_{(j,2(2k+i)+1)}$. The proof is then similar to that of Theorem 3.1. One adds and subtract $p^{(i)}(1,1)$ so that a martingale similar to (M_n) naturally appears. The limit of the remaining term is given by Lemma B.5. \square

Lemma C.8 *Under assumptions (H.0-6), for all integers $0\leq p+q+r\leq 4$, one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1}\sum_{j=1}^m\sum_{k\in\mathbb{T}_n}\delta_{(j,2k)}\delta_{(j,2k+1)}X_{(j,k)}^p\varepsilon_{(j,2k)}^q\varepsilon_{(j,2k+1)}^r=\mathbb{E}[\varepsilon_2^q\varepsilon_3^r]h^{01}(p)\mathbb{1}_{\bar{\mathcal{E}}},$$

where we used the convention $h^{01}(0)=p^{(0)}(1,1)z^0+p^{(1)}(1,1)z^1$.

Proof The proof is similar to Lemma C.4, one adds and subtracts the constant $\mathbb{E}[\varepsilon_{(j,2k)}^q\varepsilon_{(j,2k+1)}^r|\mathcal{F}_{j,\ell}^{\mathcal{O}}]$. \square

Lemma C.9 *Under assumptions (H.0-6), for all integers $1\leq p+q+r\leq 4$, one has the following a.s. limits*

$$\mathbb{1}_{\{|\mathbb{G}_n^*|>0\}}|\mathbb{T}_n^*|^{-1}\sum_{j=1}^m\sum_{k\in\mathbb{T}_n}\delta_{(j,2k)}\delta_{(j,2k+1)}X_{(j,k)}^pX_{(j,2k)}^qX_{(j,2k+1)}^r=h^{01}(p,q,r)\mathbb{1}_{\bar{\mathcal{E}}},$$

with

$$\begin{aligned} h^{01}(p,1,0) &= ah^{01}(p)+bh^{01}(p+1), & h^{01}(p,0,1) &= ch^{01}(p)+dh^{01}(p+1), \\ h^{01}(p,2,0) &= (a^2+\sigma^2)h^{01}(p)+2abh^{01}(p+1)+b^2h^{01}(p+2), \\ h^{01}(p,0,2) &= (c^2+\sigma^2)h^{01}(p)+2cdh^{01}(p+1)+d^2h^{01}(p+2), \\ h^{01}(p,3,0) &= (a^3+3a\sigma^2+\lambda)h^{01}(p)+3b(a^2+\sigma^2)h^{01}(p+1)+3ab^2h^{01}(p+2)+b^3h^{01}(p+3), \\ h^{01}(p,0,3) &= (c^3+3c\sigma^2+\lambda)h^{01}(p)+3d(c^2+\sigma^2)h^{01}(p+1)+3cd^2h^{01}(p+2)+d^3h^{01}(p+3), \\ h^{01}(p,1,1) &= (ac+\rho)h^{01}(p)+(ad+bc)h^{01}(p+1)+bdh^{01}(p+2), \\ h^{01}(p,2,1) &= ((a^2+\sigma^2)c+2a\rho+\alpha)h^{01}(p)+((a^2+\sigma^2)d+2(ac+\rho)b)h^{01}(p+1) \\ &\quad +b(2ad+bc)h^{01}(p+2)+b^2dh^{01}(p+3), \\ h^{01}(p,1,2) &= ((c^2+\sigma^2)a+2c\rho+\beta)h^{01}(p)+((c^2+\sigma^2)b+2(ac+\rho)d)h^{01}(p+1) \\ &\quad +d(ad+2bc)h^{01}(p+2)+bd^2h^{01}(p+3), \\ h^{01}(0,2,2) &= (a^2c^2+a^2\sigma^2+c^2\sigma^2+\nu^2+2a\beta+2c\alpha+4ac\rho)h^{01}(0) \\ &\quad +2(b(a(c^2+\sigma^2)+\beta+2c\rho)+d(c(a^2+\sigma^2)+\alpha+2a\rho))h^{01}(1) \\ &\quad (b^2(c^2+\sigma^2)+d^2(a^2+\sigma^2)+4bd(ac+\rho))h^{01}(2)+2bd(ad+bc)h^{01}(3)+b^2d^2h^{01}(4). \end{aligned}$$

Proof The proof is obtained by replacing $X_{(j,2k)}$ by $a+bX_k+\varepsilon_{2k}$, $X_{(j,2k)}$ by $c+dX_k+\varepsilon_{2k+1}$ and developing the exponents. One uses Lemmas C.3 and C.8 to compute the limits. \square

To conclude this section, we prove the convergence of the normalizing matrices \mathbf{S}_n^0 , \mathbf{S}_n^1 and \mathbf{S}_n^{01} where

$$\mathbf{S}_n^{01}=\sum_{j=1}^m\sum_{k\in\mathbb{T}_n}\delta_{(j,2k)}\delta_{(j,2k+1)}\begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix},$$

with the sum taken over all observed cells that have observed daughters of both types.

Lemma C.10 *Suppose that assumptions (H.0-6) are satisfied. Then, there exist definite positive matrices \mathbf{L}^0 , \mathbf{L}^1 and \mathbf{L}^{01} such that for $i \in \{0, 1\}$ one has*

$$\lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \mathbf{S}_n^i = \mathbb{1}_{\bar{\varepsilon}} \mathbf{L}^i, \quad \lim_{n \rightarrow \infty} \mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} |\mathbb{T}_n^*|^{-1} \mathbf{S}_n^{01} = \mathbb{1}_{\bar{\varepsilon}} \mathbf{L}^{01} \quad a.s.$$

where

$$\mathbf{L}^i = \begin{pmatrix} h^i(0) & h^i(1) \\ h^i(1) & h^i(2) \end{pmatrix}, \quad \mathbf{L}^{01} = \begin{pmatrix} h^{01}(0) & h^{01}(1) \\ h^{01}(1) & h^{01}(2) \end{pmatrix}.$$

Proof This is a direct consequence of Lemmas B.5 and C.3. \square

C.2 Strong consistency for the estimators of the BAR process

We could obtain the convergences of our estimators by sharp martingales results as in [3], see also B.2. However, we chose the direct approach here. Indeed, our convergences are now direct consequences of the laws of large numbers given in C.1.

Proof of Theorem 3.1, convergence of $\hat{\theta}_n$ This is a direct consequence of Lemmas C.10 and C.6. Indeed, by Lemma C.6 one has

$$\frac{\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \boldsymbol{\Sigma}_{n-1} \hat{\theta}_n = \frac{\mathbb{1}_{\{|\mathbb{G}_{n-1}^*| > 0\}}}{|\mathbb{T}_{n-1}^*|} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} X_{(j,2k)} \\ \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} \\ \delta_{(j,2k+1)} X_{(j,2k+1)} \\ \delta_{(j,2k+1)} X_{(j,k)} X_{(j,2k+1)} \end{pmatrix} \xrightarrow{n \rightarrow \infty} \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix} \theta \mathbb{1}_{\bar{\varepsilon}}.$$

And one concludes using Lemma C.10. \square

Proof of Theorem 3.1, convergence of $\hat{\sigma}_n^2$ and $\hat{\rho}_n$ This result is not as direct as the preceding one because of the presence of the $\hat{\varepsilon}_k$ in the various estimators. Take for instance the estimator $\hat{\sigma}_n^2$. For all $1 \leq j \leq m$, one has

$$\begin{aligned} \sum_{k \in \mathbb{T}_{n-1}} \hat{\varepsilon}_{(j,2k)}^2 &= \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} (X_{(j,2k)} - \hat{a}_\ell - \hat{b}_\ell X_{(j,k)})^2 \\ &= \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j,2k)} X_{(j,2k)}^2 + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \hat{a}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} + 2 \sum_{\ell=0}^{n-1} \hat{a}_\ell \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} \\ &\quad + \sum_{\ell=0}^{n-1} \hat{b}_\ell^2 \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)}^2 - 2 \sum_{\ell=0}^{n-1} \hat{a}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,2k)} - 2 \sum_{\ell=0}^{n-1} \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)}. \end{aligned}$$

Let us study the limit of the last term. One has

$$\frac{1}{\pi^n} \sum_{\ell=0}^{n-1} \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} = \frac{1}{\pi} \sum_{\ell=0}^{n-1} \frac{1}{\pi^{n-1-\ell}} \left(\hat{b}_\ell \frac{1}{\pi^\ell} \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} \right).$$

We now use Lemma B.1 with $\mathbf{A}_n = \pi^{-n}$ and $\mathbf{X}_n = \hat{b}_n \pi^{-n} \sum_{k \in \mathbb{G}_n} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)}$. We know from Lemma C.6 together with Lemma B.2 that $\pi^{-n} \sum_{k \in \mathbb{G}_n} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)}$ converges to $h^0(1, 1) W_j$, and the previous proof gives the convergence of \hat{b}_n . Thus, one obtains

$$\frac{1}{\pi^n} \sum_{\ell=0}^{n-1} \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} \xrightarrow{n \rightarrow \infty} \frac{\pi^2}{\pi - 1} W_j b h^0(1, 1),$$

so that using Lemma B.4 one has

$$\frac{\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}}}{|\mathbb{T}_n^*|} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \hat{b}_\ell \sum_{k \in \mathbb{G}_\ell} \delta_{(j,2k)} X_{(j,k)} X_{(j,2k)} \xrightarrow{n \rightarrow \infty} b h^0(1, 1) \pi^{-1}.$$

We deal with the other terms in the decomposition of the sum of $\widehat{\varepsilon}_{2k}^2$ in a similar way, using either Lemma C.3, C.5 or C.6. We can also do the same for the sums of $\widehat{\varepsilon}_{2k+1}^2$. Finally, one obtains the almost sure limit on $\overline{\mathcal{E}}$

$$\begin{aligned}\widehat{\sigma}_n^2 &\xrightarrow[n \rightarrow \infty]{} \widetilde{h}^0(2) + b^2 h^0(2) \pi^{-1} + \widetilde{h}^1(2) + d^2 h^1(2) \pi^{-1} + a^2 z^0 + c^2 z^1 + 2abh^0(1) \pi^{-1} + 2cdh^1(1) \pi^{-1} + b^2 h^0(2) \pi^{-1} \\ &\quad + d^2 h^1(2) \pi^{-1} - 2a(\widetilde{h}^0(1) + b^2 h^0(1) \pi^{-1}) - 2c(\widetilde{h}^1(1) + d^2 h^1(1) \pi^{-1}) - 2bh^0(1,1) \pi^{-1} - 2dh^1(1,1) \pi^{-1} \\ &= (z^0 + z^1) \sigma^2 = \sigma^2.\end{aligned}$$

To obtain the convergence of $\widehat{\rho}_n$ the approach is similar, using the convergence results given in Lemmas C.3, C.7, C.8 and C.9. \square

Theorem C.11 *Under assumptions (H.0-6), $\widehat{\tau}_n^4$ and $\widehat{\nu}_n^2$ converge almost surely to τ^4 and ν^2 respectively on $\overline{\mathcal{E}}$.*

Proof We work exactly along the same lines as the previous proof with higher powers. \square

C.3 Asymptotic normality for the estimators of the BAR process

We first give the asymptotic normality for $\widehat{\theta}_n$.

Proof of Theorem 3.2 for $\widehat{\theta}_n$ Define the 4×4 matrices

$$\Sigma = \begin{pmatrix} \mathbf{L}^0 & 0 \\ 0 & \mathbf{L}^1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} \sigma^2 \mathbf{L}^0 & \rho \mathbf{L}^{01} \\ \rho \mathbf{L}^{01} & \sigma^2 \mathbf{L}^1 \end{pmatrix}, \quad \Gamma_\theta = \Sigma^{-1} \Gamma \Sigma^{-1}. \quad (7)$$

We now follow the same lines as the proof of the first part of Theorem 3.2 with a different filtration. This time we use the observed sister pair-wise filtration defined as follows. For $0 \leq j \leq m$ and $p \geq 0$, let

$$\mathcal{G}_{j,p}^\mathcal{O} = \mathcal{O}_j \vee \sigma\{\delta_{(j,1)} X_{(j,1)}, (\delta_{(j,2k)} X_{(j,2k)}, \delta_{(j,2k+1)} X_{(j,2k+1)}), 1 \leq k \leq p\} \quad (8)$$

be the σ -field generated by the j -th GW tree and all the pairs of observed sister cells in genealogy j up to the daughters of cell (j, p) , and let $\mathcal{G}_p^\mathcal{O} = \bigvee_{j=1}^m \mathcal{G}_{j,p}^\mathcal{O}$ be the σ -field generated by the union of all $\mathcal{G}_{j,p}^\mathcal{O}$ for $1 \leq j \leq m$. Hence, for instance, $(\delta_{(j,2k)} \varepsilon_{(j,2k)}, \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)})$ is $\mathcal{G}_k^\mathcal{O}$ -measurable for all j . In addition, assumptions (H.1) and (H.4-5) imply that the process

$$(\delta_{(j,2k)} \varepsilon_{(j,2k)}, X_{(j,k)} \delta_{(j,2k)} \varepsilon_{(j,2k)}, \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)}, X_{(j,k)} \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)})^t$$

is a $(\mathcal{G}_k^\mathcal{O})$ -martingale difference sequence. Indeed, as the non extinction set $\overline{\mathcal{E}}$ is in $\mathcal{G}_k^\mathcal{O}$ for every $k \geq 1$, it is first easy to prove that $\mathbb{E}_{\overline{\mathcal{E}}}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}] = \mathbb{E}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}]$. Then, for $k \in \mathbb{G}_n$, using repeatedly the independence properties, one has

$$\begin{aligned}\mathbb{E}[\delta_{(j,2k)} \varepsilon_{(j,2k)} | \mathcal{G}_{k-1}^\mathcal{O}] &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{O} \vee \mathcal{F}_n \vee \sigma(\varepsilon_{j,p}, 1 \leq j \leq m, p \in \mathbb{G}_{n+1}, p \leq 2k-1)] | \mathcal{G}_{k-1}^\mathcal{O}] \\ &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_n \vee \sigma(\varepsilon_{j,p}, 1 \leq j \leq m, p \in \mathbb{G}_{n+1}, p \leq 2k-1)] | \mathcal{G}_{k-1}^\mathcal{O}] \\ &= \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_n] | \mathcal{G}_{k-1}^\mathcal{O}] = \delta_{(j,2k)} \mathbb{E}[\mathbb{E}[\varepsilon_{(j,2k)} | \mathcal{F}_{j,n}] | \mathcal{G}_{k-1}^\mathcal{O}] = 0.\end{aligned}$$

We introduce a sequence of $(\mathcal{G}_k^\mathcal{O})$ -martingales $(\mathbf{M}_p^{(n)})_{\{p \geq 1\}}$ defined for all $n, p \geq 1$ by $\mathbf{M}_p^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p \mathbf{D}_k$, with

$$\mathbf{D}_k = \sum_{j=1}^m \mathbf{D}_{(j,k)} = \sum_{j=1}^m \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ X_{(j,k)} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \\ X_{(j,k)} \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \end{pmatrix}.$$

We also introduce the sequence of stopping times $\nu_n = |\mathbb{T}_n| = 2^{n+1} - 1$. We are interested in the convergence of the process $\mathbf{M}_{\nu_n}^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^{|\mathbb{T}_n^*|} \mathbf{D}_k$. Again, it is easy to prove that

$$\mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] = \mathbb{E}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] = \sum_{j=1}^m \begin{pmatrix} \sigma^2 \varphi_{(j,k)}^0 & \rho \varphi_{(j,k)}^{01} \\ \rho \varphi_{(j,k)}^{01} & \sigma^2 \varphi_{(j,k)}^1 \end{pmatrix},$$

where for $i \in \{0, 1\}$,

$$\varphi_{(j,k)}^i = \delta_{(j,2k+i)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}, \quad \varphi_{(j,k)}^{01} = \delta_{(j,2k)} \delta_{(j,2k+1)} \begin{pmatrix} 1 & X_{(j,k)} \\ X_{(j,k)} & X_{(j,k)}^2 \end{pmatrix}.$$

Lemma C.10 yields that the $\mathbb{P}_{\bar{\mathcal{E}}}$ almost sure limit of the process $\langle \mathbf{M}^{(n)} \rangle_{\nu_n} = |\mathbb{T}_n^*|^{-1} \sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}]$ is $\mathbf{\Gamma}$, as

$$\sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}}[\mathbf{D}_k \mathbf{D}_k^t | \mathcal{G}_{k-1}^{\mathcal{O}}] = \begin{pmatrix} \sigma^2 \mathbf{S}_n^0 & \rho \mathbf{S}_n^{01} \\ \rho \mathbf{S}_n^{01} & \sigma^2 \mathbf{S}_n^1 \end{pmatrix}.$$

Therefore, the assumption A.1 of Theorem B.6 holds under $\mathbb{P}_{\bar{\mathcal{E}}}$. Thanks to assumptions (H.1) and (H.4-5) we can easily prove that for some $r > 2$, one has $\sup_{k \geq 0} \mathbb{E}[\|\mathbf{D}_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty$ a.s. which in turn implies the Lindeberg condition A.2. We can now conclude that under $\mathbb{P}_{\bar{\mathcal{E}}}$ one has

$$|\mathbb{T}_{n-1}^*|^{-1/2} \sum_{k \in \mathbb{T}_{n-1}^*} \mathbf{D}_k \xrightarrow{\mathcal{L}} \mathcal{N}(0, \mathbf{\Gamma}).$$

Finally Eq. (2) implies that $\sum_{k \in \mathbb{T}_{n-1}^*} \mathbf{D}_k = \mathbf{\Sigma}_{n-1}(\hat{\theta}_n - \theta)$. Therefore, the result is a direct consequence of Lemma C.10 together with Slutsky's Lemma. \square

We now turn to the asymptotic normality of $\hat{\sigma}_n^2$ and $\hat{\rho}_n$. The direct application of the central limit theorem for martingales to $\hat{\sigma}_n^2$ and $\hat{\rho}_n$ is not obvious because of the $\hat{\varepsilon}_{(j,2k+i)}$. We proceed along the same lines as in the proof of the convergence of $\hat{\sigma}_n^2$, using the decomposition along the generations. However, this time we need a convergence rate for $\hat{\theta}_n$ in order to apply Lemma B.1.

Theorem C.12 *Under assumptions (H.0-6), one has*

$$\mathbb{1}_{\{|\mathbb{G}_n^*| > 0\}} \|\hat{\theta}_n - \theta\|^2 = \mathcal{O}\left(\frac{\log |\mathbb{T}_{n-1}^*|}{|\mathbb{T}_{n-1}^*|}\right) \mathbb{1}_{\bar{\mathcal{E}}} \quad a.s.$$

Proof: This result is based on the asymptotic behavior of the martingale (\mathbf{M}_n) defined as follows

$$\mathbf{M}_n = \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{2j,k}, \\ \delta_{(j,2k)} X_{(j,k)} \varepsilon_{(j,2k)}, \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)}, \\ \delta_{(j,2k+1)} X_{(j,k)} \varepsilon_{(j,2k+1)} \end{pmatrix}.$$

For all $n \geq 2$, we readily deduce from the definitions of the BAR process and of our estimator $\hat{\theta}_n$ that

$$\hat{\theta}_n - \theta = \mathbf{\Sigma}_{n-1}^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \begin{pmatrix} \delta_{(j,2k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k)} X_{(j,k)} \varepsilon_{(j,2k)} \\ \delta_{(j,2k+1)} \varepsilon_{(j,2k+1)} \\ \delta_{(j,2k+1)} X_{(j,k)} \varepsilon_{(j,2k+1)} \end{pmatrix} = \mathbf{\Sigma}_{n-1}^{-1} \mathbf{M}_n.$$

The sharp asymptotic behavior of (\mathbf{M}_n) relies on properties of vector martingales. Thanks to Lemma B.4, the proof follows exactly the same lines as that of the first part of Theorem 3.2 of [3] and is not repeated here. \square

We can now turn to the end of the proof of Theorem 3.2 concerning the asymptotic normality of $\hat{\sigma}_n^2$ and $\hat{\rho}_n$.

Proof of Theorem 3.2, asymptotic normality of $\widehat{\sigma}_n^2$ Thanks to Eq. (1) and (3), we decompose $\widehat{\sigma}_n^2 - \sigma^2$ into two parts U_n and V_n

$$\begin{aligned} |\mathbb{T}_n^*|(\widehat{\sigma}_n^2 - \sigma^2) &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} (\widehat{\varepsilon}_{(j,2k)}^2 + \widehat{\varepsilon}_{(j,2k+1)}^2 - \varepsilon_{(j,2k)}^2 - \varepsilon_{(j,2k+1)}^2) \\ &\quad + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k)}(\varepsilon_{(j,2k)}^2 - \sigma^2) + \delta_{(j,2k+1)}(\varepsilon_{(j,2k+1)}^2 - \sigma^2) \\ &= \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} u_{(j,k)} + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} v_{(j,k)} = U_n + V_n, \end{aligned}$$

with

$$\begin{aligned} u_{(j,k)} &= \delta_{(j,2k)} \left((a - \widehat{a}_\ell)^2 + (b - \widehat{b}_\ell)^2 X_{(j,k)}^2 + 2(a - \widehat{a}_\ell)(b - \widehat{b}_\ell)X_{(j,k)} \right) \\ &\quad + \delta_{(j,2k+1)} \left((c - \widehat{c}_\ell)^2 + (d - \widehat{d}_\ell)^2 X_{(j,k)}^2 + 2(c - \widehat{c}_\ell)(d - \widehat{d}_\ell)X_{(j,k)} \right), \\ v_{(j,k)} &= \delta_{(j,2k)} \left(2((a - \widehat{a}_\ell) + (b - \widehat{b}_\ell)X_{(j,k)})\varepsilon_{(j,2k)} + \varepsilon_{(j,2k)}^2 - \sigma^2 \right) \\ &\quad + \delta_{(j,2k+1)} \left(2((c - \widehat{c}_\ell) + (d - \widehat{d}_\ell)X_{(j,k)})\varepsilon_{(j,2k+1)} + \varepsilon_{(j,2k+1)}^2 - \sigma^2 \right). \end{aligned}$$

We first deal with U_n and study the limit of $\pi^{-n/2}U_n$. Let us just detail the first term

$$\frac{1}{\pi^{n/2}} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k)}(a - \widehat{a}_\ell)^2 = \sum_{\ell=0}^{n-1} \pi^{(\ell-n)/2} \frac{\ell}{\pi^{\ell/2}} \frac{(a - \widehat{a}_\ell)^2}{\ell \pi^{-\ell}} \left(\frac{1}{\pi^\ell} \sum_{j=1}^m \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k)} \right) = \sum_{\ell=0}^{n-1} \pi^{(\ell-n)/2} x_\ell.$$

On the one hand, Lemmas B.5, B.3 and B.2 imply that $\pi^{-\ell} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{(j,2k)}$ converges a.s. to a finite limit. On the other hand, thanks to Theorem C.12, one has $(a - \widehat{a}_\ell)^2 (\ell \pi^{-\ell})^{-1} = \mathcal{O}(1)$ a.s. As a result, one obtains $\lim_{l \rightarrow \infty} x_l = 0$ a.s. as $\pi > 1$ by assumption. Therefore, Lemma B.1 yields

$$\lim_{n \rightarrow \infty} \frac{1}{\pi^{n/2}} \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} \delta_{j,2k} (a - \widehat{a}_\ell)^2 = 0 \quad \text{a.s.}$$

The other terms in U_n are dealt with similarly, using Lemma C.3 instead of Lemma B.5. One obtains $\lim_{n \rightarrow \infty} \pi^{-n/2}U_n = 0$ a.s. and as a result Lemma B.3 yields $\lim_{n \rightarrow \infty} |\mathbb{T}_n^*|^{-1/2}U_n = 0$. Let us now deal with the martingale term V_n . Let us remark that $|\mathbb{T}_n^*|^{-1/2}V_n = M_p^{(n)}$ with $M^{(n)} = (M_p^{(n)})_{\{p \geq 1\}}$ the sequence of $\mathcal{G}_p^{\mathcal{O}}$ -martingales defined by

$$M_p^{(n)} = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p v_k = |\mathbb{T}_n^*|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m v_{(j,k)},$$

and $\nu_n = 2^n - 1$ ($\mathcal{G}_p^{\mathcal{O}}$ defined by (8)). We want now to apply Theorem B.6 to $M^{(n)}$. Using Lemmas C.3-C.9 together with Lemma B.1 and Theorem C.12 along the same lines as above, we obtain the following limit

$$\lim_{n \rightarrow \infty} |\mathbb{T}_n^*|^{-1} \sum_{k \in \mathbb{T}_n} \mathbb{E}_{\overline{\mathcal{E}}} [v_k^2 | \mathcal{G}_{k-1}^{\mathcal{O}}] = (\tau^4 - \sigma^4) + \frac{2h^{01}(0)}{\pi} (\nu^2 - \sigma^4) \quad \mathbb{P}_{\overline{\mathcal{E}}} \text{ a.s.}$$

Therefore, assumption A.1 of Theorem B.6 holds under $\mathbb{P}_{\overline{\mathcal{E}}}$. Thanks to assumptions (H.1) and (H.4-5) we can prove that for some $r > 2$, $\sup_{k \geq 0} \mathbb{E}_{\overline{\mathcal{E}}} [\|v_k\|^r | \mathcal{G}_{k-1}^{\mathcal{O}}] < \infty$ a.s. which implies the Lindeberg condition. Therefore, we obtain that under $\mathbb{P}_{\overline{\mathcal{E}}}$

$$|\mathbb{T}_n^*|^{-1/2}V_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, (\tau^4 - \sigma^4) + \frac{2h^{01}(0)}{\pi} (\nu^2 - \sigma^4)).$$

If one sets

$$\gamma_\sigma = (\tau^4 - \sigma^4) + \frac{2h^{01}(0)}{\pi}(\nu^2 - \sigma^4), \quad (9)$$

one obtains the expected result. \square

Proof of Theorem 3.2, Asymptotic normality of $\hat{\rho}_n$. Along the same lines, we show the central limit theorem for $\hat{\rho}_n$. One has

$$|\mathbb{T}_{n-1}^{*01}|(\hat{\rho}_n - \rho) = \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} (\hat{\varepsilon}_{(j,2k)} \hat{\varepsilon}_{(j,2k+1)} - \varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)}) = \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} u'_{(j,k)} + \sum_{j=1}^m \sum_{\ell=0}^{n-1} \sum_{k \in \mathbb{G}_{\ell-1}} v'_{(j,k)} = U$$

with

$$\begin{aligned} u'_{(j,k)} &= \delta_{(j,2k)} \delta_{(j,2k+1)} \left((a - \hat{a}_l)(c - \hat{c}_l) + (b - \hat{b}_l)(d - \hat{d}_l) X_{(j,k)}^2 + ((a - \hat{a}_l)(d - \hat{d}_l) + (b - \hat{b}_l)(c - \hat{c}_l)) X_{(j,k)} \right), \\ v'_{(j,k)} &= \delta_{(j,2k)} \delta_{(j,2k+1)} \left(((a - \hat{a}_l) + (b - \hat{b}_l) X_{(j,k)}) \varepsilon_{(j,2k+1)} + ((c - \hat{c}_l) + (d - \hat{d}_l) X_{(j,k)}) \varepsilon_{(j,2k)} + \varepsilon_{(j,2k)} \varepsilon_{(j,2k+1)} \right). \end{aligned}$$

Thanks to Theorem C.12, it is easy to check that $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^{*01}|^{1/2} U'_n = 0$ a.s. Let us define a new sequence of $\mathcal{G}_p^\mathcal{O}$ -martingales $(M^{(n)})$ by

$$M_p^{(n)} = |\mathbb{T}_{n-1}^{*01}|^{-1/2} \sum_{k=1}^p v'_k = |\mathbb{T}_{n-1}^{*01}|^{-1/2} \sum_{k=1}^p \sum_{j=1}^m v'(j, k).$$

We clearly have $M_{\nu_n}^{(n)} = |\mathbb{T}_{n-1}^{*01}|^{1/2} V'_n$. We obtain the $\mathbb{P}_{\bar{\mathcal{E}}}$ - a.s. limit

$$\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^{*01}|^{-1} \sum_{k \in \mathbb{T}_n} \mathbb{E}_{\bar{\mathcal{E}}} [v_k^2 | \mathcal{G}_{k-1}^\mathcal{O}] = \nu^2 - \rho^2.$$

So we have assumption A.1 of Theorem B.6. We also derive the Lindeberg condition A.2. Consequently, we obtain that under $\mathbb{P}_{\bar{\mathcal{E}}}$, one has

$$\sqrt{|\mathbb{T}_{n-1}^{*01}|} V'_n \xrightarrow{\mathcal{L}} \mathcal{N}(0, \nu^2 - \rho^2).$$

Setting

$$\gamma_\rho = \nu^2 - \rho^2, \quad (10)$$

completes the proof of Theorem 3.2. \square

C.4 Interval estimation and tests for the BAR process

For all $n \geq 1$, define the 4×4 matrices $\hat{\Gamma}_n$ and $\hat{\Omega}_n$ by

$$\hat{\Gamma}_n = |\mathbb{T}_n^*|^{-1} \begin{pmatrix} \hat{\sigma}_n^2 \mathbf{S}_n^{\mathbf{00}} & \hat{\rho}_n \mathbf{S}_n^{\mathbf{01}} \\ \hat{\rho}_n \mathbf{S}_n^{\mathbf{01}} & \hat{\sigma}_n^2 \mathbf{S}_n^{\mathbf{11}} \end{pmatrix}, \quad \text{and} \quad \hat{\Omega}_n = \Sigma_n^{-1} \hat{\Gamma}_n \Sigma_n^{-1}.$$

Note that the matrix $\hat{\Gamma}_n$ is the empirical estimator of matrix Γ while $\hat{\Omega}_n$ is the empirical estimator of the asymptotic variance of $\hat{\theta}_n - \theta$.

Theorem C.13 *Under assumptions (H.0-6), for any $0 < \epsilon < 1$, the intervals*

$$\begin{aligned} & [\hat{a}_n - q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{1,1}; \hat{a}_n + q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{1,1}], & [\hat{b}_n - q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{2,2}; \hat{b}_n + q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{2,2}], \\ & [\hat{c}_n - q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{3,3}; \hat{c}_n + q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{3,3}], & [\hat{d}_n - q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{4,4}; \hat{d}_n + q_{1-\epsilon/2}(\hat{\Omega}_{n-1}^{1/2})_{4,4}] \end{aligned}$$

are asymptotic confidence intervals with level $1 - \epsilon$ of the parameters a, b, c and d respectively

Proof This is a straightforward consequence of the central limit Theorem 3.2 together with Stulsky's lemma as $\lim_{n \rightarrow \infty} |\mathbb{T}_{n-1}^*| \widehat{\Omega}_{n-1} = \Sigma^{-1} \Gamma \Sigma^{-1} \mathbb{P}_{\bar{\mathcal{E}}}$ a.s. thanks to Lemma C.10 and Theorem 3.1. \square

Let

$$\widehat{h}_n^{01}(0) = \widehat{p}_n^{(0)}(1, 1) |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k)} + \widehat{p}_n^{(1)}(1, 1) |\mathbb{T}_n^*|^{-1} \sum_{j=1}^m \sum_{k \in \mathbb{T}_{n-1}} \delta_{(j, 2k+1)}$$

be an empirical estimator of $h^{01}(0)$ and

$$\widehat{v}_{\sigma, n} = \frac{\widehat{\pi}_n (\widehat{\tau}_n^4 - \widehat{\sigma}_n^4) + 2\widehat{h}_n^{01}(0) (\widehat{\nu}_n^2 - \widehat{\sigma}_n^4)}{\widehat{\pi}_n}$$

be an empirical estimator of the variance term in the central limit theorem regarding σ^2 .

Theorem C.14 *Under assumptions (H.0-6), for any $0 < \epsilon < 1$, the intervals*

$$\left[\widehat{\sigma}_n^2 - q_{1-\epsilon/2} \left(\frac{\widehat{v}_{\sigma, n}}{|\mathbb{T}_n^*|} \right)^{1/2}; \widehat{\sigma}_n^2 + q_{1-\epsilon/2} \left(\frac{\widehat{v}_{\sigma, n}}{|\mathbb{T}_n^*|} \right)^{1/2} \right], \quad \left[\widehat{\rho}_n - q_{1-\epsilon/2} \left(\frac{\widehat{\nu}_n^2 - \widehat{\rho}_n^2}{|\mathbb{T}_{n-1}^{*01}|} \right)^{1/2}; \widehat{\rho}_n + q_{1-\epsilon/2} \left(\frac{\widehat{\nu}_n^2 - \widehat{\rho}_n^2}{|\mathbb{T}_{n-1}^{*01}|} \right)^{1/2} \right]$$

are asymptotic confidence intervals with level $1 - \epsilon$ of the parameters σ^2 and ρ^2 respectively.

Proof This is a again straightforward consequence of the central limit Theorem ?? together with Stulsky's lemma as

$$\lim_{n \rightarrow \infty} \widehat{v}_{\sigma, n} = \frac{\pi(\tau^4 - \sigma^4) + 2h^{01}(0)(\nu^2 - \sigma^4)}{\pi}, \quad \lim_{n \rightarrow \infty} \widehat{\nu}_n^2 - \widehat{\rho}_n^2 = \nu^2 - \rho^2,$$

$\mathbb{P}_{\bar{\mathcal{E}}}$ almost surely thanks to Lemma B.5 and Theorems 3.1 and C.11. \square

We now propose two different symmetry tests for the BAR process based on the central limit Theorem 3.2. The first one compares the couples (a, b) and (c, d) . Set

- \mathbf{H}_0^c : $(a, b) = (c, d)$ the symmetry hypothesis,
- \mathbf{H}_1^c : $(a, b) \neq (c, d)$ the alternative hypothesis.

Let $(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c$ be the test statistic defined by

$$\mathbf{Y}_n^c = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^c)^{-1/2} (\widehat{a}_n - \widehat{c}_n, \widehat{b}_n - \widehat{d}_n)^t,$$

where

$$\widehat{\Delta}_n^c = |\mathbb{T}_{n-1}^*| \mathbf{d}_{\mathbf{g}_c}^t \widehat{\Omega}_{n-1} \mathbf{d}_{\mathbf{g}_c}, \quad \mathbf{d}_{\mathbf{g}_c} = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \end{pmatrix}^t.$$

Theorem C.15 *Under assumptions (H.0-6) and the null hypothesis \mathbf{H}_0^c one has*

$$(\mathbf{Y}_n^c)^t \mathbf{Y}_n^c \xrightarrow{\mathcal{L}} \chi^2(2)$$

on $(\bar{\mathcal{E}}, \mathbb{P}_{\bar{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^c , almost surely on $\bar{\mathcal{E}}$ one has

$$\lim_{n \rightarrow \infty} \|\mathbf{Y}_n^c\|^2 = +\infty.$$

Proof We mimic again the proof of Theorem B.9 with g_c the function defined from \mathbb{R}^4 onto \mathbb{R}^2 by $g_c(x_1, x_2, x_3, x_4) = (x_1 - x_3, x_2 - x_4)^t$, so that $\mathbf{d}_{\mathbf{g}_c}$ is the gradient of g_c . \square

Our last test compares the fixed points $a/(1-b)$ and $c/(1-d)$, which are the asymptotic means of $X_{(j, 2k)}$ and $X_{(j, 2k+1)}$ respectively. Set

- \mathbf{H}_0^f : $a/(1-b) = c/(1-d)$ the symmetry hypothesis,
- \mathbf{H}_1^f : $a/(1-b) \neq c/(1-d)$ the alternative hypothesis.

Let $(Y_n^f)^2$ be the test statistic defined by

$$Y_n^f = |\mathbb{T}_{n-1}^*|^{1/2} (\widehat{\Delta}_n^f)^{-1/2} (\widehat{a}_n/(1-\widehat{b}_n) - \widehat{c}_n/(1-\widehat{d}_n)),$$

where $\widehat{\Delta}_n^f = |\mathbb{T}_{n-1}^*| \mathbf{d}_{\mathbf{g}_f}^t \widehat{\Omega}_{n-1} \mathbf{d}_{\mathbf{g}_f}$, and $\mathbf{d}_{\mathbf{g}_f} = (1/(1-\widehat{b}_n), \widehat{a}_n/(1-\widehat{b}_n)^2, -1/(1-\widehat{d}_n), -\widehat{c}_n/(1-\widehat{d}_n)^2)^t$. This test statistic has the following asymptotic properties.

Theorem C.16 *Under assumptions (H.0-6) and the null hypothesis \mathbf{H}_0^f , one has*

$$(Y_n^f)^2 \xrightarrow{\mathcal{L}} \chi^2(1)$$

on $(\overline{\mathcal{E}}, \mathbb{P}_{\overline{\mathcal{E}}})$; and under the alternative hypothesis \mathbf{H}_1^f , almost surely on $\overline{\mathcal{E}}$ one has

$$\lim_{n \rightarrow \infty} (Y_n^f)^2 = +\infty.$$

Proof We mimic one last time the proof of Theorem B.9 with g_f the function defined from \mathbb{R}^4 onto \mathbb{R} by $g_f(x_1, x_2, x_3, x_4) = (x_1/(1-x_2) - x_3/(1-x_4))$, so that $\mathbf{d}_{\mathbf{g}_f}$ is the gradient of g_f . \square

References

- [1] BERCU, B., DE SAPORTA, B., AND GÉGOUT-PETIT, A. Asymptotic analysis for bifurcating autoregressive processes via a martingale approach. *Electron. J. Probab.* 14 (2009), no. 87, 2492–2526.
- [2] COWAN, R., AND STAUDTE, R. G. The bifurcating autoregressive model in cell lineage studies. *Biometrics* 42 (1986), 769–783.
- [3] DE SAPORTA, B., GÉGOUT-PETIT, A., AND MARSALLE, L. Parameters estimation for asymmetric bifurcating autoregressive processes with missing data. *Electron. J. Statist.* 5 (2011), 1313–1353.
- [4] DE SAPORTA, B., GÉGOUT PETIT, A., AND MARSALLE, L. Asymmetry tests for bifurcating autoregressive processes with missing data. *Statistics & Probability Letters* 82, 7 (2012), 1439–1444.
- [5] DELMAS, J.-F., AND MARSALLE, L. Detection of cellular aging in a Galton-Watson process. *Stoch. Process. and Appl.* 120 (2010), 2495–2519.
- [6] DUFLO, M. *Random iterative models*, vol. 34 of *Applications of Mathematics*. Springer-Verlag, Berlin, 1997.
- [7] GUTTORP, P. *Statistical inference for branching processes*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1991. A Wiley-Interscience Publication.
- [8] GUYON, J. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *Ann. Appl. Probab.* 17, 5-6 (2007), 1538–1569.
- [9] GUYON, J., BIZE, A., PAUL, G., STEWART, E., DELMAS, J.-F., AND TADDÉI, F. Statistical study of cellular aging. In *CEMRACS 2004—mathematics and applications to biology and medicine*, vol. 14 of *ESAIM Proc.* EDP Sci., Les Ulis, 2005, pp. 100–114 (electronic).
- [10] HARRIS, T. E. *The theory of branching processes*. Die Grundlehren der Mathematischen Wissenschaften, Bd. 119. Springer-Verlag, Berlin, 1963.
- [11] HUGGINS, R. M. Robust inference for variance components models for single trees of cell lineage data. *Ann. Statist.* 24, 3 (1996), 1145–1160.

- [12] HUGGINS, R. M., AND BASAWA, I. V. Extensions of the bifurcating autoregressive model for cell lineage studies. *J. Appl. Probab.* 36, 4 (1999), 1225–1233.
- [13] HUGGINS, R. M., AND BASAWA, I. V. Inference for the extended bifurcating autoregressive model for cell lineage studies. *Aust. N. Z. J. Stat.* 42, 4 (2000), 423–432.
- [14] HUGGINS, R. M., AND STAUDTE, R. G. Variance components models for dependent cell populations. *J. AMS* 89, 425 (1994), 19–29.
- [15] MAAOUIA, F., AND TOUATI, A. Identification of multitype branching processes. *Ann. Statist.* 33, 6 (2005), 2655–2694.
- [16] STEWART, E., MADDEN, R., PAUL, G., AND TADDEI, F. Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol.* 3, 2 (2005), e45.
- [17] ZHOU, J., AND BASAWA, I. V. Maximum likelihood estimation for a first-order bifurcating autoregressive process with exponential errors. *J. Time Ser. Anal.* 26, 6 (2005), 825–842.

2 Articles relatifs au Chapitre 2

- 2.1 [dSDG10] B. DE SAPORTA, F. DUFOUR, AND K. GONZALEZ. Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.*, 20(5) :1607–1637, 2010.

NUMERICAL METHOD FOR OPTIMAL STOPPING OF PIECEWISE DETERMINISTIC MARKOV PROCESSES¹

BY BENOÎTE DE SAPORTA, FRANÇOIS DUFOUR AND KAREN GONZALEZ

Université de Bordeaux

We propose a numerical method to approximate the value function for the optimal stopping problem of a piecewise deterministic Markov process (PDMP). Our approach is based on quantization of the post jump location—inter-arrival time Markov chain naturally embedded in the PDMP, and path-adapted time discretization grids. It allows us to derive bounds for the convergence rate of the algorithm and to provide a computable ϵ -optimal stopping time. The paper is illustrated by a numerical example.

1. Introduction. The aim of this paper is to propose a computational method for optimal stopping of a piecewise deterministic Markov process $\{X(t)\}$ by using a quantization technique for an underlying discrete-time Markov chain related to the continuous-time process $\{X(t)\}$ and path-adapted time discretization grids.

Piecewise-deterministic Markov processes (PDMPs) have been introduced in the literature by Davis [6] as a general class of stochastic models. PDMPs are a family of Markov processes involving deterministic motion punctuated by random jumps. The motion of the PDMP $\{X(t)\}$ depends on three local characteristics, namely the flow ϕ , the jump rate λ and the transition measure Q , which specifies the post-jump location. Starting from x the motion of the process follows the flow $\phi(x, t)$ until the first jump time T_1 which occurs either spontaneously in a Poisson-like fashion with rate $\lambda(\phi(x, t))$ or when the flow $\phi(x, t)$ hits the boundary of the state-space. In either case the location of the process at the jump time $T_1 : X(T_1) = Z_1$ is selected by the transition measure $Q(\phi(x, T_1), \cdot)$. Starting from Z_1 , we now select the next interjump time $T_2 - T_1$ and postjump location $X(T_2) = Z_2$. This gives a piecewise deterministic trajectory for $\{X(t)\}$ with jump times $\{T_k\}$ and post jump locations $\{Z_k\}$ which follows the flow ϕ between two jumps. A suitable choice of the state space and the local characteristics ϕ , λ and Q provide stochastic models covering a great number of problems of operations research [6].

Optimal stopping problems have been studied for PDMPs in [3, 5, 6, 9, 11, 13]. In [11] the author defines an operator related to the first jump time of the process

Received March 2009; revised September 2009.

¹Supported by ARPEGE program of the French National Agency of Research (ANR), project “FAUTOCOES,” number ANR-09-SEGI-004.

AMS 2000 subject classifications. Primary 93E20; secondary 93E03, 60J25.

Key words and phrases. Optimal stopping, piecewise deterministic Markov processes, quantization, numerical method, dynamic programming.

and shows that the value function of the optimal stopping problem is a fixed point for this operator. The basic assumption in this case is that the final cost function is continuous along trajectories, and it is shown that the value function will also have this property. In [9, 13] the authors adopt some stronger continuity assumptions and boundary conditions to show that the value function of the optimal stopping problem satisfies some variational inequalities related to integro-differential equations. In [6], Davis assumes that the value function is bounded and locally Lipschitz along trajectories to show that the variational inequalities are necessary and sufficient to characterize the value function of the optimal stopping problem. In [5], the authors weakened the continuity assumptions of [6, 9, 13]. A paper related to our work is [3] by Costa and Davis. It is the only one presenting a computational technique for solving the optimal stopping problem for a PDMP based on a discretization of the state space similar to the one proposed by Kushner in [12]. In particular, the authors in [3] derive a convergence result for the approximation scheme but no estimation of the rate of convergence is derived.

Quantization methods have been developed recently in numerical probability, nonlinear filtering or optimal stochastic control with applications in finance [1, 2, 14–17]. More specifically, powerful and interesting methods have been developed in [1, 2, 17] for computing the Snell-envelope associated to discrete-time Markov chains and diffusion processes. Roughly speaking, the approach developed in [1, 2, 17] for studying the optimal stopping problem for a continuous-time diffusion process $\{Y(t)\}$ is based on a time-discretization scheme to obtain a discrete-time Markov chain $\{\bar{Y}_k\}$. It is shown that the original continuous-time optimization problem can be converted to an auxiliary optimal stopping problem associated with the discrete-time Markov chain $\{\bar{Y}_k\}$. Under some suitable assumptions, a rate of convergence of the auxiliary value function to the original one can be derived. Then, in order to address the optimal stopping problem of the discrete-time Markov chain, a twofold computational method is proposed. The first step consists in approximating the Markov chain by a quantized process. There exists an extensive literature on quantization methods for random variables and processes. We do not pretend to present here an exhaustive panorama of these methods. However, the interested reader may, for instance, consult [10, 14, 17] and the references therein. The second step is to approximate the conditional expectations which are used to compute the backward dynamic programming formula by the conditional expectation related to the quantized process. This procedure leads to a tractable formula called a *quantization tree algorithm* (see Proposition 4 in [1] or Section 4.1 in [17]). Providing the cost function and the Markov kernel are Lipschitz, some bounds and rates of convergence are obtained (see, e.g., Section 2.2.2 in [1]).

As regards PDMPs, it was shown in [11] that the value function of the optimal stopping problem can be calculated by iterating a functional operator, labeled L [see (3.5) for its definition], which is related to a continuous-time maximization and a discrete-time dynamic programming formula. Thus, in order to approximate the value function of the optimal stopping problem of a PDMP $\{X(t)\}$, a natural

approach would have been to follow the same lines as in [1, 2, 17]. However, their method cannot be directly applied to our problem for two main reasons related to the specificities of PDMPs.

First, PDMPs are in essence discontinuous at random times. Therefore, as pointed out in [11], it will be problematic to convert the original optimization problem into an optimal stopping problem associated to a time discretization of $\{X(t)\}$ with nice convergence properties. In particular, it appears ill-advised to propose as in [1] a fixed-step time-discretization scheme $\{X(k\Delta)\}$ of the original process $\{X(t)\}$. Besides, another important intricacy concerns the transition semi-group $\{P_t\}_{t \in \mathbb{R}_+}$ of $\{X(t)\}$. On the one hand, it cannot be explicitly calculated from the local characteristics (ϕ, λ, Q) of the PDMP (see [4, 7]). Consequently, it will be complicated to express the Markov kernel P_Δ associated with the Markov chain $\{X(k\Delta)\}$. On the other hand, the Markov chain $\{X(k\Delta)\}$ is, in general, not even a Feller chain (see [6], pages 76 and 77), and therefore it will be hard to ensure it is K -Lipschitz (see Definition 1 in [1]).

Second, the other main difference stems from the fact that the function appearing in the backward dynamic programming formula associated with L and the reward function g is not continuous even if some strong regularity assumptions are made on g . Consequently, the approach developed in [1, 2, 17] has to be refined since it can only handle conditional expectations of Lipschitz-continuous functions.

However, by using the special structure of PDMPs, we are able to overcome both these obstacles. Indeed, associated to the PDMP $\{X(t)\}$, there exists a natural embedded discrete-time Markov chain $\{\Theta_k\}$ with $\Theta_k = (Z_k, S_k)$ where S_k is given by the inter-arrival time $T_k - T_{k-1}$. The main operator L can be expressed using the chain $\{\Theta_k\}$ and a continuous-time maximization. We first convert the continuous-time maximization of operator L into a discrete-time maximization by using a path-dependent time-discretization scheme. This enables us to approximate the value function by the solution of a backward dynamic programming equation in discrete-time involving conditional expectation of the Markov chain $\{\Theta_k\}$. Then, a natural approximation of this optimization problem is obtained by replacing $\{\Theta_k\}$ by its quantized approximation. It must be pointed out that this optimization problem is related to the calculation of conditional expectations of indicator functions of the Markov chain $\{\Theta_k\}$. As said above, it is not straightforward to obtain convergence results as in [1, 2, 17]. We deal successfully with indicator functions by showing that the event on which the discontinuity actually occurs is of small enough probability. This enables us to provide a rate of convergence for the approximation scheme.

In addition, and more importantly, this numerical approximation scheme enables us to propose a computable stopping rule which also is an ϵ -optimal stopping time of the original stopping problem. Indeed, for any $\epsilon > 0$ one can construct a stopping time, labeled τ , such that

$$V(x) - \epsilon \leq \mathbf{E}_x[g(X(\tau))] \leq V(x),$$

where $V(x)$ is the optimal value function associated to the original stopping problem. Our computational approach is attractive in the sense that it does not require any additional calculations. Moreover, we can characterize how far it is from optimal in terms of the value function. In [1], Section 2.2.3, Proposition 6, another criteria for the approximation of the optimal stopping time has been proposed. In the context of PDMPs, it must be noticed that an optimal stopping time does not generally exist as shown in [11], Section 2.

An additional result extends Theorem 1 of Gugerli [11] by showing that the iteration of operator L provides a sequence of random variables which corresponds to a *quasi-Snell* envelope associated with the reward process $\{g(X(t))\}_{t \in \mathbb{R}_+}$ where the horizon time is random and given by the jump times $(T_n)_{n \in \{0, \dots, N\}}$ of the process $\{X(t)\}_{t \in \mathbb{R}_+}$.

The paper is organized as follows. In Section 2 we give a precise definition of PDMPs and state our notation and assumptions. In Section 3, we state the optimal stopping problem, recall and refine some results from [11]. In Section 4, we build an approximation of the value function. In Section 5, we compute the error between the approximate value function and the real one. In Section 6 we propose a computable ϵ -optimal stopping time and evaluate its sharpness. Finally in Section 7 we present a numerical example. Technical results are postponed to the [Appendix](#).

2. Definitions and assumptions. We first give a precise definition of a piecewise deterministic Markov process. Some general assumptions are presented in the second part of this section. Let us introduce first some standard notation. Let M be a metric space. $\mathbf{B}(M)$ is the set of real-valued, bounded, measurable functions defined on M . The Borel σ -field of M is denoted by $\mathcal{B}(M)$. Let Q be a Markov kernel on $(M, \mathcal{B}(M))$ and $w \in \mathbf{B}(M)$, $Qw(x) = \int_M w(y) Q(x, dy)$ for $x \in M$. For $(a, b) \in \mathbb{R}^2$, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

2.1. Definition of a PDMP. Let E be an open subset of \mathbb{R}^n , ∂E its boundary and \overline{E} its closure. A PDMP is determined by its local characteristics (ϕ, λ, Q) where:

- The flow $\phi : \mathbb{R}^n \times \mathbb{R} \rightarrow \mathbb{R}^n$ is a one-parameter group of homeomorphisms: ϕ is continuous, $\phi(\cdot, t)$ is a homeomorphism for each $t \in \mathbb{R}$ satisfying $\phi(\cdot, t + s) = \phi(\phi(\cdot, s), t)$.

For all x in E , let us denote

$$t^*(x) \doteq \inf\{t > 0 : \phi(x, t) \in \partial E\}$$

with the convention $\inf \emptyset = \infty$.

- The jump rate $\lambda : \overline{E} \rightarrow \mathbb{R}_+$ is assumed to be a measurable function satisfying

$$(\forall x \in E), \quad (\exists \varepsilon > 0) \quad \text{such that } \int_0^\varepsilon \lambda(\phi(x, s)) ds < \infty.$$

- Q is a Markov kernel on $(\bar{E}, \mathcal{B}(\bar{E}))$ satisfying the following property:

$$(\forall x \in \bar{E}), \quad Q(x, E - \{x\}) = 1.$$

From these characteristics, it can be shown [6], pages 62–66, that there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \{\mathbf{P}_x\}_{x \in E})$ such that the motion of the process $\{X(t)\}$ starting from a point $x \in E$ may be constructed as follows. Take a random variable T_1 such that

$$\mathbf{P}_x(T_1 > t) \doteq \begin{cases} e^{-\Lambda(x,t)}, & \text{for } t < t^*(x), \\ 0, & \text{for } t \geq t^*(x), \end{cases}$$

where for $x \in E$ and $t \in [0, t^*(x)]$

$$\Lambda(x, t) \doteq \int_0^t \lambda(\phi(x, s)) ds.$$

If T_1 generated according to the above probability is equal to infinity, then for $t \in \mathbb{R}_+$, $X(t) = \phi(x, t)$. Otherwise select independently an E -valued random variable (labelled Z_1) having distribution $Q(\phi(x, T_1), \cdot)$, namely $\mathbf{P}_x(Z_1 \in A) = Q(\phi(x, T_1), A)$ for any $A \in \mathcal{B}(\bar{E})$. The trajectory of $\{X(t)\}$ starting at x , for $t \leq T_1$, is given by

$$X(t) \doteq \begin{cases} \phi(x, t), & \text{for } t < T_1, \\ Z_1, & \text{for } t = T_1. \end{cases}$$

Starting from $X(T_1) = Z_1$, we now select the next inter-jump time $T_2 - T_1$ and post-jump location $X(T_2) = Z_2$ in a similar way.

This gives a strong Markov process $\{X(t)\}$ with jump times $\{T_k\}_{k \in \mathbb{N}}$ (where $T_0 = 0$). Associated with $\{X(t)\}$, there exists a discrete time process $(\Theta_n)_{n \in \mathbb{N}}$ defined by $\Theta_n = (Z_n, S_n)$ with $Z_n = X(T_n)$ and $S_n = T_n - T_{n-1}$ for $n \geq 1$ and $S_0 = 0$. Clearly, the process $(\Theta_n)_{n \in \mathbb{N}}$ is a Markov chain.

We introduce a standard assumption (see, e.g., equations (24.4) or (24.8) in [6]).

ASSUMPTION 2.1. For all $(x, t) \in E \times \mathbb{R}_+$, $\mathbf{E}_x[\sum_k \mathbf{1}_{\{T_k \leq t\}}] < \infty$.

In particular, it implies that $T_k \rightarrow \infty$ as $k \rightarrow \infty$.

For $n \in \mathbb{N}$, let \mathcal{M}_n be the family of all $\{\mathcal{F}_t\}$ -stopping times which are dominated by T_n , and for $n < p$, let $\mathcal{M}_{n,p}$ be the family of all $\{\mathcal{F}_t\}$ -stopping times ν satisfying $T_n \leq \nu \leq T_p$. Let \mathbf{B}^c denote the set of all real-valued, bounded, measurable functions, w defined on \bar{E} and continuous along trajectories up to the jump time horizon: for any $x \in E$, $w(\phi(x, \cdot))$ is continuous on $[0, t^*(x)]$. Let \mathbf{L}^c be the set of all real-valued, bounded, measurable functions, w defined on \bar{E} and Lipschitz along trajectories:

1. there exists $[w]_1 \in \mathbb{R}_+$ such that for any $(x, y) \in E^2$, $u \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|w(\phi(x, u)) - w(\phi(y, u))| \leq [w]_1 |x - y|;$$

2. there exists $[w]_2 \in \mathbb{R}_+$ such that for any $x \in E$, and $(t, s) \in [0, t^*(x)]^2$, one has

$$|w(\phi(x, t)) - w(\phi(x, s))| \leq [w]_2 |t - s|;$$

3. there exists $[w]_* \in \mathbb{R}_+$ such that for any $(x, y) \in E^2$, one has

$$|w(\phi(x, t^*(x))) - w(\phi(y, t^*(y)))| \leq [w]_* |x - y|.$$

In the sequel, for any function f in \mathbf{B}^c , we denote by C_f its bound

$$C_f = \sup_{x \in E} |f(x)|,$$

and for any Lipschitz-continuous function f in $\mathbf{B}(E)$ or $\mathbf{B}(\overline{E})$, we denote by $[f]$ its Lipschitz constant

$$[f] = \sup_{x \neq y \in E} \frac{|f(x) - f(y)|}{|x - y|}.$$

REMARK 2.2. \mathbf{L}^c is a subset of \mathbf{B}^c and any function in \mathbf{L}^c is Lipschitz on \overline{E} with $[w] \leq [w]_1$.

Finally, as a convenient abbreviation, we set for any $x \in \overline{E}$, $\lambda Qw(x) = \lambda(x)Qw(x)$.

2.2. *Assumptions.* The following assumptions will be in force throughout.

ASSUMPTION 2.3. The jump rate λ is bounded and there exists $[\lambda]_1 \in \mathbb{R}_+$ such that for any $(x, y) \in E^2$, $u \in [0, t^*(x) \wedge t^*(y)]$,

$$|\lambda(\phi(x, u)) - \lambda(\phi(y, u))| \leq [\lambda]_1 |x - y|.$$

ASSUMPTION 2.4. The exit time t^* is bounded and Lipschitz-continuous on E .

ASSUMPTION 2.5. The Markov kernel Q is Lipschitz in the following sense: there exists $[Q] \in \mathbb{R}_+$ such that for any function $w \in \mathbf{L}^c$ the following two conditions are satisfied:

1. for any $(x, y) \in E^2$, $u \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|Qw(\phi(x, u)) - Qw(\phi(y, u))| \leq [Q][w]_1 |x - y|;$$

2. for any $(x, y) \in E^2$, one has

$$|Qw(\phi(x, t^*(x))) - Qw(\phi(y, t^*(y)))| \leq [Q][w]_* |x - y|.$$

The reward function g associated with the optimal stopping problem satisfies the following hypothesis.

ASSUMPTION 2.6. g is in \mathbf{L}^c .

3. Optimal stopping problem. From now on, assume that the distribution of $X(0)$ is given by δ_{x_0} for a fixed state $x_0 \in E$. Let us consider the following optimal stopping problem for a fixed integer N :

$$(3.1) \quad \sup_{\tau \in \mathcal{M}_N} \mathbb{E}_{x_0}[g(X(\tau))].$$

This problem has been studied by Gugerli [11].

Note that Assumption 2.3 yields $\Lambda(x, t) < \infty$ for all x, t . Hence, for all x in E , the jump time horizon $s^*(x)$ defined in [11] by $t^*(x) \wedge \inf\{t \geq 0, e^{-\Lambda(x,t)} = 0\}$ is equal to the exit time $t^*(x)$. Therefore, operators $H: \mathbf{B}(\bar{E}) \rightarrow \mathbf{B}(E \times \mathbb{R}_+)$, $I: \mathbf{B}(E) \rightarrow \mathbf{B}(E \times \mathbb{R}_+)$, $J: \mathbf{B}(E) \times \mathbf{B}(\bar{E}) \rightarrow \mathbf{B}(E \times \mathbb{R}_+)$, $K: \mathbf{B}(E) \rightarrow \mathbf{B}(E)$ and $L: \mathbf{B}(E) \times \mathbf{B}^c \rightarrow \mathbf{B}^c$ introduced by Gugerli ([11], Section 2) reduce to

$$(3.2) \quad \begin{aligned} Hf(x, t) &= f(\phi(x, t \wedge t^*(x)))e^{-\Lambda(x, t \wedge t^*(x))}, \\ Iw(x, t) &= \int_0^{t \wedge t^*(x)} \lambda Qw(\phi(x, s))e^{-\Lambda(x, s)} ds, \\ J(w, f)(x, t) &= Iw(x, t) + Hf(x, t), \end{aligned}$$

$$(3.3) \quad \begin{aligned} Kw(x) &= \int_0^{t^*(x)} \lambda Qw(\phi(x, s))e^{-\Lambda(x, s)} ds \\ &\quad + Qw(\phi(x, t^*(x)))e^{-\Lambda(x, t^*(x))}, \end{aligned}$$

$$L(w, h)(x) = \sup_{t \geq 0} J(w, h)(x, t) \vee Kw(x).$$

It is easy to derive a probabilistic interpretation of operators H, I, K and L in terms of the embedded Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$.

LEMMA 3.1. For all $x \in E$, $w \in \mathbf{B}(E)$, $f \in \mathbf{B}(\bar{E})$, $h \in \mathbf{B}^c$ and $t \geq 0$, one has

$$(3.4) \quad \begin{aligned} Hf(x, t) &= f(\phi(x, t \wedge t^*(x)))\mathbf{P}_x(S_1 \geq t \wedge t^*(x)), \\ Iw(x, t) &= \mathbf{E}_x[w(Z_1)\mathbf{1}_{\{S_1 < t \wedge t^*(x)\}}], \\ Kw(x) &= \mathbf{E}_x[w(Z_1)], \end{aligned}$$

$$(3.5) \quad \begin{aligned} L(w, h)(x) &= \sup_{u \leq t^*(x)} \{\mathbf{E}_x[w(Z_1)\mathbf{1}_{\{S_1 < u\}}] + h(\phi(x, u))\mathbf{P}_x(S_1 \geq u)\} \\ &\quad \vee \mathbf{E}_x[w(Z_1)]. \end{aligned}$$

For a reward function $g \in \mathbf{B}^c$, it has been shown in [11] that the value function can be recursively constructed by the following procedure:

$$\sup_{\tau \in \mathcal{M}_N} \mathbb{E}_{x_0}[g(X(\tau))] = v_0(x_0)$$

with

$$\begin{cases} v_N = g, \\ v_k = L(v_{k+1}, g), \quad \text{for } k \leq N - 1. \end{cases}$$

DEFINITION 3.2. Introduce the random variables $(V_n)_{n \in \{0, \dots, N\}}$ by

$$V_n = v_n(Z_n)$$

or equivalently

$$(3.6) \quad \begin{aligned} V_n = & \sup_{u \leq t^*(Z_n)} \{ \mathbf{E}[v_{n+1}(Z_{n+1}) \mathbf{1}_{\{S_{n+1} < u\}} + g(\phi(Z_n, u)) \mathbf{1}_{\{S_{n+1} \geq u\}} | Z_n] \} \\ & \vee \mathbf{E}[v_{n+1}(Z_{n+1}) | Z_n]. \end{aligned}$$

The following result shows that the sequence $(V_n)_{n \in \{0, \dots, N\}}$ corresponds to a *quasi-Snell envelope* associated with the reward process $\{g(X(t))\}_{t \in \mathbb{R}_+}$ where the horizon time is random and given by the jump times $(T_n)_{n \in \{0, \dots, N\}}$ of the process $\{X(t)\}_{t \in \mathbb{R}_+}$:

THEOREM 3.3. Consider an integer $n < N$. Then

$$V_n = \sup_{\nu \in \mathcal{M}_{n,N}} \mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}].$$

PROOF. Let $\nu \in \mathcal{M}_{n,N}$. According to Proposition B.4 and Corollary B.6 in Appendix B, there exists $\hat{\nu}: E \times (\mathbb{R}_+ \times E)^n \times \Omega \rightarrow \mathbb{R}_+$ such that for all $(z_0, \gamma) \in E \times (\mathbb{R}_+ \times E)^n$ the mapping $\hat{\nu}(z_0, \gamma): \Omega \rightarrow \mathbb{R}_+$ is an $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ -stopping time satisfying $\hat{\nu}(z_0, \gamma) \leq T_{N-n}$, and $\nu = T_n + \hat{\nu}(Z_0, \Gamma_n, \theta_{T_n})$, where $\Gamma_n = (S_1, Z_1, \dots, S_n, Z_n)$ and θ is the shift operator. For $(z_0, \gamma) \in E \times (\mathbb{R}_+ \times E)^n$ define $\mathcal{W}: E \times (\mathbb{R}_+ \times E)^n \rightarrow \mathbb{R}$ by

$$\mathcal{W}(z_0, \gamma) = \mathbf{E}_{z_n}[g(X(\hat{\nu}(z_0, \gamma)))] \leq \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))],$$

where $\gamma = (s_1, z_1, \dots, s_n, z_n)$. Hence, the strong Markov property of the process $\{X(t)\}$ yields

$$\mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}] = \mathbf{E}_{x_0}[g(X(T_n + \hat{\nu}(Z_0, \Gamma_n, \theta_{T_n}))) | \mathcal{F}_{T_n}] = \mathcal{W}(Z_0, \Gamma_n).$$

Consequently, one has

$$\mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}] \leq \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))]$$

and, therefore, one has

$$(3.7) \quad \sup_{\nu \in \mathcal{M}_{n,N}} \mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}] \leq \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))].$$

Conversely, consider $\tau \in \mathcal{M}_{N-n}$. It is easy to show that $T_n + \tau \circ \theta_{T_n} \in \mathcal{M}_{n,N}$. The strong Markov property of the process $\{X(t)\}$ again yields

$$\mathbf{E}_{Z_n}[g(X(\tau))] = \mathbf{E}_{x_0}[g(X(T_n + \tau \circ \theta_{T_n})) | \mathcal{F}_{T_n}] \leq \sup_{\nu \in \mathcal{M}_{n,N}} \mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}]$$

and hence we obtain

$$(3.8) \quad \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))] \leq \sup_{\nu \in \mathcal{M}_{n,N}} \mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}].$$

Combining equations (3.7) and (3.8), one has

$$\sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))] = \sup_{\nu \in \mathcal{M}_{n,N}} \mathbf{E}_{x_0}[g(X(\nu)) | \mathcal{F}_{T_n}].$$

Finally, it is proved in [11], Theorem 1, that $v_n(x) = \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_x[g(X(\tau))]$, whence

$$V_n = \sup_{\tau \in \mathcal{M}_{N-n}} \mathbf{E}_{Z_n}[g(X(\tau))],$$

showing the result. \square

4. Approximation of the value function. To approximate the sequence of value functions (V_n) , we proceed in two steps. First, the continuous-time maximization of operator L is converted into a discrete-time maximization by using a path-dependent time-discretization scheme to give a new operator L^d . In particular, it is important to remark that these time-discretization grids depend on the the post-jump locations $\{Z_k\}$ of the PDMP (see Definition 4.1 and Remark 4.2). Second, the conditional expectations of the Markov chain (Θ_k) in the definition of L^d are replaced by the conditional expectations of its quantized approximation $(\hat{\Theta}_k)$ to define an operator \hat{L}^d .

First, we define the path-adapted discretization grids as follows.

DEFINITION 4.1. For $z \in E$, set $\Delta(z) \in]0, t^*(z)[$. Define $n(z) = \text{int}(\frac{t^*(z)}{\Delta(z)}) - 1$, where $\text{int}(x)$ denotes the greatest integer smaller than or equal to x . The set of points $(t_i)_{i \in \{0, \dots, n(z)\}}$ with $t_i = i \Delta(z)$ is denoted by $G(z)$. This is the grid associated with the time interval $[0, t^*(z)]$.

REMARK 4.2. It is important to note that, for all $z \in E$, not only one has $t^*(z) \notin G(z)$, but also $\max G(z) = t_{n(z)} \leq t^*(z) - \Delta(z)$. This property is crucial for the sequel.

DEFINITION 4.3. Consider for $w \in \mathbf{B}(E)$ and $z \in E$,

$$L^d(w, g)(z) = \max_{s \in G(z)} \{ \mathbf{E}[w(Z_1) \mathbf{1}_{\{S_1 < s\}} + g(\phi(z, s)) \mathbf{1}_{\{S_1 \geq s\}} | Z_0 = z] \} \\ \vee \mathbf{E}[w(Z_1) | Z_0 = z].$$

Now let us turn to the quantization of (Θ_n) . The quantization algorithm will provide us with a finite grid $\Gamma_n^\Theta \subset E \times \mathbb{R}_+$ at each time $0 \leq n \leq N$ as well as weights for each point of the grid (see, e.g., [1, 14, 17]). Set $p \geq 1$ such that Θ_n has finite

moments at least up to the order p and let p_n be the closest-neighbor projection from $E \times \mathbb{R}_+$ onto Γ_n^\ominus (for the distance of norm p ; if there are several equally close neighbors, pick the one with the smallest index). Then the quantization of Θ_n is defined by

$$\widehat{\Theta}_n = (\widehat{Z}_n, \widehat{S}_n) = p_n(Z_n, S_n).$$

We will also denote by Γ_n^Z , the projection of Γ_n^\ominus on E , and by Γ_n^S , the projection of Γ_n^\ominus on \mathbb{R}_+ .

In practice, one will first compute the quantization grids and weights, and then compute a path-adapted time-grid for each $z \in \Gamma_n^Z$, for all $0 \leq n \leq N - 1$. Hence, there is only a finite number of time grids to compute, and like the quantization grids, they can be computed and stored off-line.

The definition of the discretized operators now naturally follows the characterization given in Lemma 3.1.

DEFINITION 4.4. For $k \in \{1, \dots, N\}$, $w \in \mathbf{B}(\Gamma_k^Z)$, $z \in \Gamma_{k-1}^Z$, and $s \in \mathbb{R}_+$

$$\begin{aligned} \widehat{J}_k(w, g)(z, s) &= \mathbf{E}[w(\widehat{Z}_k)\mathbf{1}_{\{\widehat{S}_k < s\}} + g(\phi(z, s))\mathbf{1}_{\{\widehat{S}_k \geq s\}} | \widehat{Z}_{k-1} = z], \\ \widehat{K}_k(w)(z) &= \mathbf{E}[w(\widehat{Z}_k) | \widehat{Z}_{k-1} = z], \\ \widehat{L}_k^d(w, g)(z) &= \max_{s \in G(z)} \{\widehat{J}_k(w, g)(z, s)\} \vee \widehat{K}_k(w)(z). \end{aligned}$$

Note that $\widehat{\Theta}_n$ is a random variable taking finitely many values, hence the expectations above actually are finite sums, the probability of each atom being given by its weight on the quantization grid. We can now give the complete construction of the sequence approximating (V_n) .

DEFINITION 4.5. Consider $\widehat{v}_N(z) = g(z)$ where $z \in \Gamma_N^Z$ and for $k \in \{1, \dots, N\}$

$$(4.1) \quad \widehat{v}_{k-1}(z) = \widehat{L}_k^d(\widehat{v}_k, g)(z),$$

where $z \in \Gamma_{k-1}^Z$.

DEFINITION 4.6. The approximation of V_k is denoted by

$$(4.2) \quad \widehat{V}_k = \widehat{v}_k(\widehat{Z}_k)$$

for $k \in \{0, \dots, N\}$.

5. Error estimation for the value function. We are now able to state our main result, namely the convergence of our approximation scheme with an upper bound for the rate of convergence.

THEOREM 5.1. *Set $n \in \{0, \dots, N - 1\}$, and suppose that $\Delta(z)$, for $z \in \Gamma_n^z$, are chosen such that*

$$\min_{z \in \Gamma_n^z} \{\Delta(z)\} > (2C_\lambda)^{-1/2} ([t^*] \|\widehat{Z}_n - Z_n\|_p + \|S_{n+1} - \widehat{S}_{n+1}\|_p)^{1/2}.$$

Then the discretization error for V_n is no greater than the following:

$$\begin{aligned} \|V_n - \widehat{V}_n\|_p &\leq \|V_{n+1} - \widehat{V}_{n+1}\|_p + \alpha \|\Delta(\widehat{Z}_n)\|_p + \beta_n \|\widehat{Z}_n - Z_n\|_p \\ &\quad + 2[v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p \\ &\quad + \gamma ([t^*] \|\widehat{Z}_n - Z_n\|_p + \|S_{n+1} - \widehat{S}_{n+1}\|_p)^{1/2}, \end{aligned}$$

where $\alpha = [g]_2 + 2C_g C_\lambda$, $\beta_n = [v_n] + [v_{n+1}]_1 E_2 + C_g E_4 + ([g]_1 + [g]_2 [t^]) \vee ([v_{n+1}]_* [Q])$, $\gamma = 4C_g (2C_\lambda)^{1/2}$, and E_2 and E_4 are defined in Appendix A.*

Recall that $V_N = g(Z_N)$ and $\widehat{V}_N = g(\widehat{Z}_N)$, hence $\|V_N - \widehat{V}_N\|_p \leq [g] \|\widehat{Z}_N - Z_N\|_p$. In addition, the quantization error $\|\Theta_n - \widehat{\Theta}_n\|_p$ goes to zero as the number of points in the grids goes to infinity (see, e.g., [14]). Hence $|V_0 - \widehat{V}_0|$ can be made arbitrarily small by an adequate choice of the discretizations parameters.

Remark that the square root in the last error term is the price to pay for integrating noncontinuous functions, see the definition of operator J with the indicator functions, and the introduction of Section 5.2.

To prove Theorem 5.1, we split the left-hand side difference into four terms

$$\|V_n - \widehat{V}_n\|_p \leq \sum_{i=1}^4 \Xi_i,$$

where

$$\begin{aligned} \Xi_1 &= \|v_n(Z_n) - v_n(\widehat{Z}_n)\|_p, \\ \Xi_2 &= \|L(v_{n+1}, g)(\widehat{Z}_n) - L^d(v_{n+1}, g)(\widehat{Z}_n)\|_p, \\ \Xi_3 &= \|L^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n)\|_p, \\ \Xi_4 &= \|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{Z}_n)\|_p. \end{aligned}$$

The first term is easy enough to handle thanks to Proposition A.7 in Appendix A.2.

LEMMA 5.2. *A upper bound for Ξ_1 is*

$$\|v_n(Z_n) - v_n(\widehat{Z}_n)\|_p \leq [v_n] \|Z_n - \widehat{Z}_n\|_p.$$

We are going to study the other terms one by one in the following sections.

5.1. Second term. In this part we study the error induced by the replacement of the supremum over all nonnegative t smaller than or equal to $t^*(z)$ by the maximum over the finite grid $G(z)$ in the definition of operator L .

LEMMA 5.3. *Let $w \in \mathbf{L}^c$. Then for all $z \in E$,*

$$\left| \sup_{t \leq t^*(z)} J(w, g)(z, t) - \max_{s \in G(z)} J(w, g)(z, s) \right| \leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) \Delta(z).$$

PROOF. Clearly, there exists $\bar{t} \in [0, t^*(z)]$ such that $\sup_{t \leq t^*(z)} J(w, g)(z, t) = J(w, g)(z, \bar{t})$, and there exists $0 \leq i \leq n(z)$ such that $\bar{t} \in [t_i, t_{i+1}]$ [with $t_{n(z)+1} = t^*(z)$]. Consequently, Lemma A.2 yields

$$\begin{aligned} 0 &\leq \sup_{t \leq t^*(z)} J(w, g)(z, t) - \max_{s \in G(z)} J(w, g)(z, s) \\ &\leq J(w, g)(z, \bar{t}) - J(w, g)(z, t_i) \\ &\leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) |\bar{t} - t_i| \\ &\leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) |t_{i+1} - t_i|, \end{aligned}$$

implying the result. \square

Turning back to the second error term, one gets the following bound.

LEMMA 5.4. *A upper bound for Ξ_2 is*

$$\|L(v_{n+1}, g)(\widehat{Z}_n) - L^d(v_{n+1}, g)(\widehat{Z}_n)\|_p \leq ([g]_2 + 2C_g C_\lambda) \|\Delta(\widehat{Z}_n)\|_p.$$

PROOF. From the definition of L and L^d we readily obtain

$$\begin{aligned} &\|L(v_{n+1}, g)(\widehat{Z}_n) - L^d(v_{n+1}, g)(\widehat{Z}_n)\|_p \\ &\leq \left\| \sup_{t \leq t^*(\widehat{Z}_n)} J(v_{n+1}, g)(\widehat{Z}_n, t) - \max_{s \in G(\widehat{Z}_n)} J(v_{n+1}, g)(\widehat{Z}_n, s) \right\|_p. \end{aligned}$$

Now in view of the previous lemma, one has

$$\begin{aligned} &\|L(v_{n+1}, g)(\widehat{Z}_n) - L^d(v_{n+1}, g)(\widehat{Z}_n)\|_p \\ &\leq (C_{v_{n+1}} C_\lambda + [g]_2 + C_g C_\lambda) \|\Delta(\widehat{Z}_n)\|_p. \end{aligned}$$

Finally, note that $C_{v_{n+1}} = C_g$ (see Appendix A.2), completing the proof. \square

5.2. *Third term.* This is the crucial part of our derivation, where we need to compare conditional expectations relative to the real Markov chain (Z_n, S_n) and its quantized approximation $(\widehat{Z}_n, \widehat{S}_n)$. The main difficulty stems from the fact that some functions inside the expectations are indicator functions and in particular they are not Lipschitz-continuous. We manage to overcome this difficulty by proving that the event on which the discontinuity actually occurs is of small enough probability; this is the aim of the following two lemmas.

LEMMA 5.5. For all $n \in \{0, \dots, N-1\}$ and $0 < \eta < \min_{z \in \Gamma_n^Z} \{\Delta(z)\}$,

$$\begin{aligned} & \left\| \max_{s \in G(\widehat{Z}_n)} \mathbf{E}[\mathbf{1}_{\{S_{n+1} < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < s\}} | \widehat{Z}_n] \right\|_p \\ & \leq \frac{2}{\eta} \|S_{n+1} - \widehat{S}_{n+1}\|_p + C_\lambda \eta + \frac{2[t^*] \|Z_n - \widehat{Z}_n\|_p}{\eta}. \end{aligned}$$

PROOF. Set $0 < \eta < \min_{z \in \Gamma_n^Z} \{\Delta(z)\}$. Remark that the difference of indicator functions is nonzero if and only if S_{n+1} and \widehat{S}_{n+1} are on either side of s . Hence, one has

$$|\mathbf{1}_{\{S_{n+1} < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < s\}}| \leq \mathbf{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta/2\}} + \mathbf{1}_{\{|S_{n+1} - s| \leq \eta/2\}}.$$

This yields

$$\begin{aligned} (5.1) \quad & \left\| \max_{s \in G(\widehat{Z}_n)} \mathbf{E}[\mathbf{1}_{\{S_{n+1} < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < s\}} | \widehat{Z}_n] \right\|_p \\ & \leq \left\| \mathbf{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta/2\}} \right\|_p + \left\| \max_{s \in G(\widehat{Z}_n)} \mathbf{E}[\mathbf{1}_{\{s - \eta/2 \leq S_{n+1} \leq s + \eta/2\}} | \widehat{Z}_n] \right\|_p. \end{aligned}$$

On the one hand, Chebyshev's inequality yields

$$(5.2) \quad \left\| \mathbf{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta/2\}} \right\|_p^p = \mathbf{P}\left(|S_{n+1} - \widehat{S}_{n+1}| > \frac{\eta}{2}\right) \leq \frac{2^p \|S_{n+1} - \widehat{S}_{n+1}\|_p^p}{\eta^p}.$$

On the other hand, as $s \in G(\widehat{Z}_n)$ and by definition of η , one has $s + \eta < t^*(\widehat{Z}_n)$ (see Remark 4.2). Thus, one has

$$\begin{aligned} (5.3) \quad & \mathbf{E}[\mathbf{1}_{\{s - \eta/2 \leq S_{n+1} \leq s + \eta/2\}} | \widehat{Z}_n] \\ & = \mathbf{E}[\mathbf{E}[\mathbf{1}_{\{s - \eta/2 \leq S_{n+1} \leq s + \eta/2\}} | Z_n] | \widehat{Z}_n] \\ & \leq \mathbf{E}\left[\int_{s - \eta/2}^{s + \eta/2} \lambda(\phi(Z_n, u)) du | \widehat{Z}_n\right] + \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) \leq s + \eta/2\}} | \widehat{Z}_n] \\ & \leq \eta C_\lambda + \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) \leq t^*(\widehat{Z}_n) - \eta/2\}} | \widehat{Z}_n]. \end{aligned}$$

Combining equations (5.1)–(5.3), the result follows. \square

LEMMA 5.6. For all $n \in \{0, \dots, N\}$ and $0 < \eta < \min_{z \in \Gamma_n^Z} \{\Delta(z)\}$,

$$\left\| \mathbf{1}_{\{t^*(Z_n) < t^*(\widehat{Z}_n) - \eta\}} \right\|_p \leq \frac{[t^*] \|Z_n - \widehat{Z}_n\|_p}{\eta}.$$

PROOF. We use Chebyshev's inequality again. One clearly has

$$\begin{aligned} \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) < t^*(\widehat{Z}_n) - \eta\}}]^p & = \mathbf{P}(t^*(Z_n) < t^*(\widehat{Z}_n) - \eta) \\ & \leq \mathbf{P}(|t^*(Z_k) - t^*(\widehat{Z}_k)| > \eta) \\ & \leq \frac{[t^*]^p \|Z_k - \widehat{Z}_k\|_p^p}{\eta^p}, \end{aligned}$$

showing the result. \square

Now we turn to the consequences of replacing the Markov chain (Z_n, S_n) by its quantized approximation $(\widehat{Z}_n, \widehat{S}_n)$ in the conditional expectations.

LEMMA 5.7. *Let $w \in \mathbf{L}^c$, then one has*

$$\begin{aligned} & |\mathbf{E}[w(Z_{n+1})|Z_n = \widehat{Z}_n] - \mathbf{E}[w(\widehat{Z}_{n+1})|\widehat{Z}_n]| \\ & \leq (C_w E_4 + [w]_1 E_2 + [w]_*[Q])\mathbf{E}[|Z_n - \widehat{Z}_n||\widehat{Z}_n] \\ & \quad + [w]\mathbf{E}[|Z_{n+1} - \widehat{Z}_{n+1}||\widehat{Z}_n]. \end{aligned}$$

PROOF. First, note that

$$\begin{aligned} & \mathbf{E}[w(Z_{n+1})|Z_n = \widehat{Z}_n] - \mathbf{E}[w(\widehat{Z}_{n+1})|\widehat{Z}_n] \\ & = \mathbf{E}[w(Z_{n+1})|Z_n = \widehat{Z}_n] - \mathbf{E}[w(Z_{n+1})|\widehat{Z}_n] \\ & \quad + \mathbf{E}[w(Z_{n+1})|\widehat{Z}_n] - \mathbf{E}[w(\widehat{Z}_{n+1})|\widehat{Z}_n]. \end{aligned}$$

On the one hand, Remark 2.2 yields

$$|\mathbf{E}[w(Z_{n+1})|\widehat{Z}_n] - \mathbf{E}[w(\widehat{Z}_{n+1})|\widehat{Z}_n]| \leq [w]\mathbf{E}[|Z_{n+1} - \widehat{Z}_{n+1}||\widehat{Z}_n].$$

On the other hand, recall that by construction of the quantized process, one has $(\widehat{Z}_n, \widehat{S}_n) = p_n(Z_n, S_n)$. Hence we have the following property: $\sigma\{\widehat{Z}_n\} \subset \sigma\{Z_n, S_n\}$. By using the special structure of the PDMP $\{X(t)\}$, we have $\sigma\{Z_n, S_n\} \subset \mathcal{F}_{T_n}$. Now, by using the Markov property of the process $\{X(t)\}$, it follows that

$$\mathbf{E}[w(Z_{n+1})|\widehat{Z}_n] = \mathbf{E}[\mathbf{E}[w(Z_{n+1})|\mathcal{F}_{T_n}||\widehat{Z}_n] = \mathbf{E}[\mathbf{E}[w(Z_{n+1})|Z_n||\widehat{Z}_n].$$

Equation (3.4) thus yields

$$\begin{aligned} & \mathbf{E}[w(Z_{n+1})|Z_n = \widehat{Z}_n] - \mathbf{E}[w(Z_{n+1})|\widehat{Z}_n] \\ & = \mathbf{E}[\mathbf{E}[w(Z_{n+1})|Z_n = \widehat{Z}_n] - \mathbf{E}[w(Z_{n+1})|Z_n||\widehat{Z}_n] \\ & = \mathbf{E}[Kw(\widehat{Z}_n) - Kw(Z_n)|\widehat{Z}_n]. \end{aligned}$$

Now we use Lemma A.4 to conclude. \square

Now we combine the preceding lemmas to derive the third error term.

LEMMA 5.8. *For all $0 < \eta < \min_{z \in \Gamma_n^z} \{\Delta(z)\}$, an upper bound for Ξ_3 is*

$$\begin{aligned} & \|L^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n)\|_p \\ & \leq \left\{ [v_{n+1}]_1 E_2 + C_g E_4 + 2C_g \frac{[t^*]}{\eta} \right. \\ & \quad \left. + ([g]_1 + [g]_2[t^*]) \vee ([v_{n+1}]_*[Q]) \right\} \|\widehat{Z}_n - Z_n\|_p \\ & \quad + [v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p + 2C_g \left(2C_\lambda \eta + \frac{\|S_{n+1} - \widehat{S}_{n+1}\|_p}{\eta} \right). \end{aligned}$$

PROOF. To simplify notation, set $\Psi(x, y, t) = v_{n+1}(y)\mathbf{1}_{\{t < s\}} + g(\phi(x, t)) \times \mathbf{1}_{\{t \geq s\}}$. From the definition of L^d and \widehat{L}_{n+1}^d , one readily obtains

$$\begin{aligned}
 & |L^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n)| \\
 & \leq \max_{s \in G(\widehat{Z}_n)} |\mathbf{E}[\Psi(Z_n, Z_{n+1}, S_{n+1}) | Z_n = \widehat{Z}_n] \\
 (5.4) \quad & \quad - \mathbf{E}[\Psi(\widehat{Z}_n, \widehat{Z}_{n+1}, \widehat{S}_{n+1}) | \widehat{Z}_n]| \\
 & \quad \vee |\mathbf{E}[v_{n+1}(Z_{n+1}) | Z_n = \widehat{Z}_n] - \mathbf{E}[v_{n+1}(\widehat{Z}_{n+1}) | \widehat{Z}_n]|.
 \end{aligned}$$

On the one hand, combining Lemma 5.7 and the fact that v_{n+1} is in \mathbf{L}^c (see Proposition A.7), we obtain

$$\begin{aligned}
 & |\mathbf{E}[v_{n+1}(Z_{n+1}) | Z_n = \widehat{Z}_n] - \mathbf{E}[v_{n+1}(\widehat{Z}_{n+1}) | \widehat{Z}_n]| \\
 (5.5) \quad & \leq [v_{n+1}] \mathbf{E}[|Z_{n+1} - \widehat{Z}_{n+1}| | \widehat{Z}_n] \\
 & \quad + (C_g E_4 + [v_{n+1}]_1 E_2 + [v_{n+1}]_* [Q]) \mathbf{E}[|Z_n - \widehat{Z}_n| | \widehat{Z}_n].
 \end{aligned}$$

On the other hand, similar arguments as in the proof of Lemma 5.7 yield

$$\begin{aligned}
 & \mathbf{E}[\Psi(Z_n, Z_{n+1}, S_{n+1}) | Z_n = \widehat{Z}_n] - \mathbf{E}[\Psi(\widehat{Z}_n, \widehat{Z}_{n+1}, \widehat{S}_{n+1}) | \widehat{Z}_n] \\
 & = \mathbf{E}[\mathbf{E}[\Psi(Z_n, Z_{n+1}, S_{n+1}) | Z_n = \widehat{Z}_n] \\
 (5.6) \quad & \quad - \mathbf{E}[\Psi(Z_n, Z_{n+1}, S_{n+1}) | Z_n = Z_n] | \widehat{Z}_n] \\
 & \quad + \mathbf{E}[\Psi(Z_n, Z_{n+1}, S_{n+1}) | \widehat{Z}_n] - \mathbf{E}[\Psi(\widehat{Z}_n, \widehat{Z}_{n+1}, \widehat{S}_{n+1}) | \widehat{Z}_n] \\
 & = \Upsilon_1 + \Upsilon_2.
 \end{aligned}$$

The second difference of the right-hand side of (5.6), labeled Υ_2 , clearly satisfies

$$\begin{aligned}
 & |\Upsilon_2| \leq [v_{n+1}] \mathbf{E}[|\widehat{Z}_{n+1} - Z_{n+1}| | \widehat{Z}_n] + [g]_1 \mathbf{E}[|\widehat{Z}_n - Z_n| | \widehat{Z}_n] \\
 (5.7) \quad & \quad + 2C_g \mathbf{E}[|\mathbf{1}_{\{S_{n+1} < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < s\}}| | \widehat{Z}_n].
 \end{aligned}$$

Let us turn now to the first difference of the right-hand side of (5.6), labeled Υ_1 . We meet another difficulty here. Indeed, we know by construction that $s < t^*(\widehat{Z}_n)$, but we know nothing regarding the relative positions of s and $t^*(Z_n)$. In the event where $s \leq t^*(Z_n)$ as well, we recognize operator J inside the expectations. In the opposite event $s > t^*(Z_n)$, we crudely bound Ψ by $C_{v_{n+1}} + C_g = 2C_g$. Hence, one obtains

$$\begin{aligned}
 & |\Upsilon_1| \leq \mathbf{E}[|J(v_{n+1}, g)(\widehat{Z}_n, s) - J(v_{n+1}, g)(Z_n, s)| \mathbf{1}_{\{s \leq t^*(Z_n)\}} | \widehat{Z}_n] \\
 & \quad + 2C_g \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) < s\}} | \widehat{Z}_n].
 \end{aligned}$$

Now Lemma A.3 gives an upper bound for the first term. As for the indicator function, by definition of $G(\widehat{Z}_n)$ and our choice of η , we have $s < t^*(\widehat{Z}_n) - \eta$.

Thus, one has

$$(5.8) \quad |\Upsilon_1| \leq (C_g E_1 + [v_{n+1}]_1 E_2 + E_3) \mathbf{E}[|\widehat{Z}_n - Z_n| | \widehat{Z}_n] \\ + 2C_g \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) < t^*(\widehat{Z}_n) - \eta\}} | \widehat{Z}_n].$$

Now, combining (5.4), (5.5), (5.7) and (5.8), and the fact that $C_g E_1 + E_3 = C_g E_4 + [g]_1 + [g]_2 [t^*]$, one gets

$$|L^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n)| \\ \leq \{[v_{n+1}]_1 E_2 + C_g E_4 \\ + ([g]_1 + [g]_2 [t^*]) \vee ([v_{n+1}]_* [Q])\} \mathbf{E}[|\widehat{Z}_n - Z_n| | \widehat{Z}_n] \\ + [v_{n+1}] \mathbf{E}[|\widehat{Z}_{n+1} - Z_{n+1}| | \widehat{Z}_n] \\ + 2C_g \mathbf{E}[\mathbf{1}_{\{t^*(Z_n) < t^*(\widehat{Z}_n) - \eta\}} | \widehat{Z}_n] \\ + 2C_g \max_{s \in G(\widehat{Z}_n)} \mathbf{E}[\mathbf{1}_{\{S_{n+1} < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < s\}} | \widehat{Z}_n].$$

Finally, we conclude by taking the L^p norm on both sides and using Lemmas 5.5 and 5.6. \square

5.3. *Fourth term.* The last error term is a mere comparison of two finite sums.

LEMMA 5.9. *An upper bound for \mathfrak{E}_4 is*

$$\|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{Z}_n)\|_p \\ \leq [v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p.$$

PROOF. By definition of operator \widehat{L}_n^d , one has

$$\|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{Z}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{Z}_n)\|_p \\ = \left\| \max_{s \in G(\widehat{Z}_n)} \{ \mathbf{E}[v_{n+1}(\widehat{Z}_{n+1}) \mathbf{1}_{\{\widehat{S}_{n+1} < s\}} + g(\phi(\widehat{Z}_n, s)) \mathbf{1}_{\{\widehat{S}_{n+1} \geq s\}} | \widehat{Z}_n] \} \right. \\ \left. \vee \mathbf{E}[v_{n+1}(\widehat{Z}_{n+1}) | \widehat{Z}_n] \right. \\ \left. - \max_{s \in G(\widehat{Z}_n)} \{ \mathbf{E}[\widehat{v}_{n+1}(\widehat{Z}_{n+1}) \mathbf{1}_{\{\widehat{S}_{n+1} < s\}} \right. \\ \left. + g(\phi(\widehat{Z}_n, s)) \mathbf{1}_{\{\widehat{S}_{n+1} \geq s\}} | \widehat{Z}_n] \} \vee \mathbf{E}[\widehat{v}_{n+1}(\widehat{Z}_{n+1}) | \widehat{Z}_n] \right\|_p \\ \leq \|\mathbf{E}[v_{n+1}(\widehat{Z}_{n+1}) - \widehat{v}_{n+1}(\widehat{Z}_{n+1}) | \widehat{Z}_n]\|_p \\ \leq \|v_{n+1}(\widehat{Z}_{n+1}) - v_{n+1}(Z_{n+1})\|_p + \|v_{n+1}(Z_{n+1}) - \widehat{v}_{n+1}(\widehat{Z}_{n+1})\|_p.$$

We conclude using the fact that $v_{n+1} \in \mathbf{L}^c$ (see Proposition A.7) and the definitions of V_n and \widehat{V}_n . \square

5.4. *Proof of Theorem 5.1.* We can finally turn to the proof of Theorem 5.1. Lemmas 5.2, 5.4, 5.8 and 5.9 from the preceding sections directly yield, for all $0 < \eta < \min_{z \in \Gamma_n^z} \{\Delta(z)\}$,

$$\begin{aligned} \|V_n - \widehat{V}_n\|_p &\leq [v_n] \|\widehat{Z}_n - Z_n\|_p + ([g]_2 + 2C_g C_\lambda) \|\Delta(\widehat{Z}_n)\|_p \\ &\quad + \left\{ [v_{n+1}]_1 E_2 + C_g E_4 + 2C_g \frac{[t^*]}{\eta} \right. \\ &\quad \left. + ([g]_1 + [g]_2 [t^*]) \vee ([v_{n+1}]_* [Q]) \right\} \|\widehat{Z}_n - Z_n\|_p \\ &\quad + [v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p + 2C_g \left(2C_\lambda \eta + \frac{\|S_{n+1} - \widehat{S}_{n+1}\|_p}{\eta} \right) \\ &\quad + [v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p. \end{aligned}$$

The optimal choice for η clearly satisfies

$$2C_\lambda \eta = \frac{1}{\eta} ([t^*] \|\widehat{Z}_n - Z_n\|_p + \|S_{n+1} - \widehat{S}_{n+1}\|_p),$$

providing it also satisfies the condition $0 < \eta < \min_{z \in \Gamma_n^z} \{\Delta(z)\}$. Hence, rearranging the terms above, one gets the expected result

$$\begin{aligned} \|V_n - \widehat{V}_n\|_p &\leq \|V_{n+1} - \widehat{V}_{n+1}\|_p + ([g]_2 + 2C_g C_\lambda) \|\Delta(\widehat{Z}_n)\|_p \\ &\quad + \{[v_n] + [v_{n+1}]_1 E_2 + C_g E_4 \\ &\quad + ([g]_1 + [g]_2 [t^*]) \vee ([v_{n+1}]_* [Q])\} \|\widehat{Z}_n - Z_n\|_p \\ &\quad + 2[v_{n+1}] \|\widehat{Z}_{n+1} - Z_{n+1}\|_p \\ &\quad + 4C_g (2C_\lambda)^{1/2} ([t^*] \|\widehat{Z}_n - Z_n\|_p + \|S_{n+1} - \widehat{S}_{n+1}\|_p)^{1/2}. \end{aligned}$$

6. Numerical construction of an ϵ -optimal stopping time. In [11], Theorem 1, Gugerli defined an ϵ -optimal stopping time for the original problem. Roughly speaking, this stopping time depends on the embedded Markov chain (Θ_n) and on the optimal value function. Therefore, a natural candidate for an ϵ -optimal stopping time should be obtained by replacing the Markov chain (Θ_n) and the optimal value function by their *quantized approximations*. However, this leads to un-tractable comparisons between some quantities involving (Θ_n) and its quantized approximation. It is then far from obvious to show that this method would provide a computable ϵ -optimal stopping rule. Nonetheless, by modifying the approach of Gugerli [11], we are able to propose a numerical construction of an ϵ -optimal stopping time of the original stopping problem.

Here is how we proceed. First, recall that p_n be the closest-neighbor projection from $E \times \mathbb{R}_+$ onto Γ_n^Θ , and for all $(z, s) \in E \times \mathbb{R}_+$ define $(\widehat{z}_n, \widehat{s}_n) = p_n(z, s)$. Note

that \widehat{z}_n and \widehat{s}_n depend on both z and s . Now, for $n \in \{1, \dots, N\}$, define

$$s_n^*(z, s) = \min \left\{ t \in G(\widehat{z}_{n-1}) \mid \widehat{J}_n(\widehat{v}_n, g)(\widehat{z}_{n-1}, t) = \max_{u \in G(\widehat{z}_{n-1})} \widehat{J}_n(\widehat{v}_n, g)(\widehat{z}_{n-1}, u) \right\}$$

and

$$r_{n,\beta}(z, s) = \begin{cases} t^*(z), & \text{if } \widehat{K}_n \widehat{v}_n(\widehat{z}_{n-1}) > \max_{u \in G(\widehat{z}_{n-1})} \widehat{J}_n(\widehat{v}_n, g)(\widehat{z}_{n-1}, u), \\ s_n^*(z, s) \mathbf{1}_{\{s_n^*(z,s) < t^*(z)\}} + (t^*(z) - \beta) \mathbf{1}_{\{s_n^*(z,s) \geq t^*(z)\}}, & \text{otherwise.} \end{cases}$$

Note the use of both the real jump time horizon $t^*(z)$ and the quantized approximations of K , J and (z, s) . Set

$$\tau_1 = r_{N,\beta}(Z_0, S_0) \wedge T_1$$

and for $n \in \{1, \dots, N - 1\}$, set

$$\tau_{n+1} = \begin{cases} r_{N-n,\beta}(Z_0, S_0), & \text{if } T_1 > r_{N-n,\beta}(Z_0, S_0), \\ T_1 + \tau_n \circ \theta_{T_1}, & \text{otherwise.} \end{cases}$$

Our stopping rule is then defined by τ_N .

REMARK 6.1. This procedure is especially appealing because it requires no more calculation: we have already computed the values of \widehat{K}_n and \widehat{J}_n on the grids. One just has to store the point where the maximum of \widehat{J}_n is reached.

LEMMA 6.2. τ_N is an $\{\mathcal{F}_T\}$ -stopping time.

PROOF. Set $U_1 = r_{1,\beta}(Z_0, S_0)$ and for $2 \leq k \leq N$ $U_k = r_{k,\beta}(Z_{k-1}, S_{k-1}) \times \mathbf{1}_{\{r_{k-1,\beta}(Z_{k-2}, S_{k-2}) \geq S_{k-1}\}}$. One then clearly has $\tau_N = \sum_{k=1}^N U_k \wedge S_k$ which is an $\{\mathcal{F}_T\}$ -stopping time by Proposition B.5. \square

Now let us show that this stopping time provides a good approximation of the value function V_0 . Namely, for all $z \in E$ set

$$\bar{v}_n(z) = \mathbf{E}[g(X_{\tau_{N-n}}) \mid Z_n = z]$$

and in accordance to our previous notation introduce, for $n \in \{1, \dots, N - 1\}$

$$\bar{V}_n = \bar{v}_n(Z_n).$$

The comparison between V_0 and \bar{V}_0 is provided by the next two theorems.

THEOREM 6.3. Set $n \in \{0, \dots, N - 2\}$ and suppose the discretization parameters are chosen such that there exists $0 < a < 1$ satisfying

$$\frac{\beta}{a} = (2C_\lambda)^{-1/2} \left(\frac{[t^*]}{1-a} \|\widehat{Z}_n - Z_n\|_p + \|S_{n+1} - \widehat{S}_{n+1}\|_p \right)^{1/2} < \min_{z \in \Gamma_n^z} \{\Delta(z)\}.$$

Then one has

$$\begin{aligned} \|\bar{V}_n - V_n\|_p &\leq \|\bar{V}_{n+1} - V_{n+1}\|_p + \|\hat{V}_{n+1} - V_{n+1}\|_p + \|\hat{V}_n - V_n\|_p \\ &\quad + 2[v_{n+1}]\|Z_{n+1} - \hat{Z}_{n+1}\|_p + a_n\|Z_n - \hat{Z}_n\|_p \\ &\quad + 4C_g(2C_\lambda)^{1/2} \left(\frac{[t^*]}{1-a} \|\hat{Z}_n - Z_n\|_p + \|S_{n+1} - \hat{S}_{n+1}\|_p \right)^{1/2} \end{aligned}$$

with $a_n = (2[v_{n+1}]_1 E_2 + 2C_g C_{t^*}[\lambda]_1 (2 + C_{t^*} C_\lambda) + (4C_g C_\lambda [t^*] + 2[v_{n+1}]_* [Q]) \vee (3[g]_1))$.

PROOF. The definition of τ_n and the strong Markov property of the process $\{X(t)\}$ yield

$$\begin{aligned} \bar{v}_n(Z_n) &= \mathbf{E}[g(X_{r_{n+1,\beta}(Z_n, S_n)}) \mathbf{1}_{\{S_{n+1} > r_{n+1,\beta}(Z_n, S_n)\}} | Z_n] \\ &\quad + \mathbf{E}[\bar{v}_{n+1}(Z_{n+1}) \mathbf{1}_{\{S_{n+1} \leq r_{n+1,\beta}(Z_n, S_n)\}} | Z_n] \\ &= \mathbf{1}_{\{r_{n+1,\beta}(Z_n, S_n) \geq t^*(Z_n)\}} K \bar{v}_{n+1}(Z_n) \\ &\quad + \mathbf{1}_{\{r_{n+1,\beta}(Z_n, S_n) < t^*(Z_n)\}} J(\bar{v}_{n+1}, g)(Z_n, r_{n+1,\beta}(Z_n, S_n)). \end{aligned}$$

However, our definition of $r_{n,\beta}$ with the special use of parameter β implies

$$\{r_{n+1,\beta}(Z_n, S_n) \geq t^*(Z_n)\} = \left\{ \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n) > \max_{s \in G(\hat{Z}_n)} \hat{J}_{n+1}(\hat{v}_{n+1}, g)(\hat{Z}_n, s) \right\}.$$

Consequently, one obtains

$$\begin{aligned} \bar{v}_n(Z_n) &= \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n) \vee \max_{s \in G(\hat{Z}_n)} \hat{J}_{n+1}(\hat{v}_{n+1}, g)(\hat{Z}_n, s) \\ &\quad + \mathbf{1}_{\{r_{n+1,\beta}(Z_n, S_n) \geq t^*(Z_n)\}} [K \bar{v}_{n+1}(Z_n) - \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n)] \\ (6.1) \quad &\quad + \mathbf{1}_{\{r_{n+1,\beta}(Z_n, S_n) < t^*(Z_n)\}} \left[J(\bar{v}_{n+1}, g)(Z_n, r_{n+1,\beta}(Z_n, S_n)) \right. \\ &\quad \left. - \max_{s \in G(\hat{Z}_n)} \hat{J}_{n+1}(\hat{v}_{n+1}, g)(\hat{Z}_n, s) \right]. \end{aligned}$$

Let us study the term with operator K . First, we insert V_n to be able to use our work from the previous section (we cannot directly apply it to \bar{v}_n because it may not be Lipschitz-continuous). Clearly, one has

$$\begin{aligned} &|K \bar{v}_{n+1}(Z_n) - \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n)| \\ (6.2) \quad &\leq \mathbf{E}[|\bar{V}_{n+1} - V_{n+1}| | Z_n] + |K v_{n+1}(Z_n) - \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n)|. \end{aligned}$$

Similar calculations to those of Lemmas A.4, 5.7 and 5.9, and equation (5.5) yield

$$\begin{aligned} &|K v_{n+1}(Z_n) - \hat{K}_{n+1} \hat{v}_{n+1}(\hat{Z}_n)| \\ (6.3) \quad &\leq (C_g E_4 + [v_{n+1}]_1 E_2 + [v_{n+1}]_* [Q]) (|Z_n - \hat{Z}_n| + \mathbf{E}[|Z_n - \hat{Z}_n| | \hat{Z}_n]) \\ &\quad + 2[v_{n+1}] \mathbf{E}[|Z_{n+1} - \hat{Z}_{n+1}| | \hat{Z}_n] + \mathbf{E}[|V_{n+1} - \hat{V}_{n+1}| | \hat{Z}_n]. \end{aligned}$$

Now we turn to operator J . Set $R_n = r_{n+1,\beta}(Z_n, S_n)$. We first study the case when $R_n = s_{n+1}^*(Z_n, S_n) < t^*(Z_n)$. By definition, one has

$$\widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, R_n) = \max_{s \in G(\widehat{Z}_n)} \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, s).$$

As above, we insert V_n and obtain

$$(6.4) \quad \left| \left[J(\bar{v}_{n+1}, g)(Z_n, R_n) - \max_{s \in G(\widehat{Z}_n)} \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, s) \right] \mathbf{1}_{\{R_n = s_{n+1}^*(Z_n, S_n)\}} \right| \\ \leq \mathbf{E}[|\bar{V}_{n+1} - V_{n+1}| | Z_n] \mathbf{1}_{\{R_n = s_{n+1}^*(Z_n, S_n)\}} \\ + |J(v_{n+1}, g)(Z_n, R_n) - \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, R_n)| \mathbf{1}_{\{R_n = s_{n+1}^*(Z_n, S_n)\}}.$$

Again, similar arguments as those used for Lemmas A.3, 5.6 and 5.9, and equations (5.6), (5.7) and (5.8) yield, on $\{R_n = s_{n+1}^*(Z_n, S_n)\}$

$$(6.5) \quad |J(v_{n+1}, g)(Z_n, R_n) - \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, R_n)| \\ \leq ([v_{n+1}]_1 E_2 + [g]_1 + C_g C_{t^*}[\lambda]_1 (2 + C_{t^*} C_\lambda)) \\ \times (|Z_n - \widehat{Z}_n| + \mathbf{E}[|Z_n - \widehat{Z}_n| | \widehat{Z}_n]) \\ + 2[v_{n+1}] \mathbf{E}[|Z_{n+1} - \widehat{Z}_{n+1}| | \widehat{Z}_n] + \mathbf{E}[|V_{n+1} - \widehat{V}_{n+1}| | \widehat{Z}_n] \\ + [g]_1 \mathbf{E}[|Z_n - \widehat{Z}_n| | \widehat{Z}_n] \\ + 2C_g \mathbf{E}[|\mathbf{1}_{\{S_{n+1} < R_n\}} - \mathbf{1}_{\{\widehat{S}_{n+1} < R_n\}}| | \widehat{Z}_n].$$

Note that all the constants with a factor $[t^*]$ have vanished because we know here that both $R_n < t^*(Z_n)$ and $R_n < t^*(\widehat{Z}_n)$ hold on $\{R_n = s_{n+1}^*(Z_n, S_n)\}$.

Finally, on $\{s^*(Z_n) \geq t^*(Z_n) = R_n + \beta\}$, by construction of the grid $G(\widehat{Z}_n)$ (see Remark 4.2), one has for all $0 < \eta < \min_{z \in \Gamma_n^Z} \{\Delta(z)\}$,

$$R_n = t^*(Z_n) - \beta < s^*(Z_n) < t^*(\widehat{Z}_n) - \eta.$$

Consequently, using the crude bound

$$|J(\bar{v}_{n+1}, g)(Z_n, R_n)| + \left| \max_{s \in G(\widehat{Z}_n)} \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, s) \right| \leq 2C_g,$$

one obtains

$$(6.6) \quad \left| J(\bar{v}_{n+1}, g)(Z_n, r_{n+1,\beta}(Z_n, S_n)) - \max_{s \in G(\widehat{Z}_n)} \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{Z}_n, s) \right| \\ \times \mathbf{1}_{\{r_{n+1,\beta}(Z_n, S_n) = t^*(Z_n) - \beta\}} \\ \leq 2C_g |\mathbf{1}_{\{t^*(Z_n) - \beta < t^*(\widehat{Z}_n) - \eta\}}|.$$

Now the combination of equations (6.1)–(6.6) and Lemmas 5.5 and 5.6 yields, for all $\beta < \eta < \min_{z \in \Gamma_n^z} \{\Delta(z)\}$

$$\begin{aligned} \|\bar{V}_n - \hat{V}_n\|_p &\leq \|\bar{V}_{n+1} - V_{n+1}\|_p + \|V_{n+1} - \hat{V}_{n+1}\|_p + 2[v_{n+1}]\|Z_{n+1} - \hat{Z}_{n+1}\|_p \\ &\quad + \|Z_n - \hat{Z}_n\|_p (2[v_{n+1}]_1 E_2 + 2C_g C_{t^*}[\lambda]_1 (2 + C_{t^*} C_\lambda) \\ &\quad \quad \quad + (4C_g C_\lambda [t^*] + 2[v_{n+1}]_* [Q]) \vee (3[g]_1)) \\ &\quad + 2C_g \left(2C_\lambda \eta + \frac{1}{\eta} \|S_{n+1} - \hat{S}_{n+1}\|_p + \frac{[t^*]}{\eta - \beta} \|Z_n - \hat{Z}_n\|_p \right). \end{aligned}$$

Now suppose there exists $0 < a < 1$ such that $\eta = a^{-1}\beta$. Then the optimal choice for η satisfies

$$2C_\lambda \eta = \frac{1}{\eta} \left(\frac{[t^*]}{1-a} \|\hat{Z}_n - Z_n\|_p + \|S_{n+1} - \hat{S}_{n+1}\|_p \right),$$

providing it also satisfies the condition $0 < \eta < \min_{z \in \Gamma_n^z} \{\Delta(z)\}$, hence the result. \square

Theorem 6.3 gives a recursive error estimation. Here is the initializing step.

THEOREM 6.4. *Suppose the discretization parameters are chosen such that there exists $0 < a < 1$ satisfying*

$$\frac{\beta}{a} = (2C_\lambda)^{-1/2} \left(\frac{[t^*]}{1-a} \|\hat{Z}_{N-1} - Z_{N-1}\|_p + \|S_N - \hat{S}_N\|_p \right)^{1/2} < \min_{z \in \Gamma_{N-1}^z} \{\Delta(z)\}.$$

Then one has

$$\begin{aligned} &\|\bar{V}_{N-1} - V_{N-1}\|_p \\ &\leq \|\hat{V}_{N-1} - V_{N-1}\|_p + 3[g]\|Z_N - \hat{Z}_N\|_p + a_{N-1}\|Z_{N-1} - \hat{Z}_{N-1}\|_p \\ &\quad + 4C_g (2C_\lambda)^{1/2} \left(\frac{[t^*]}{1-a} \|\hat{Z}_{N-1} - Z_{N-1}\|_p + \|S_N - \hat{S}_N\|_p \right)^{1/2} \end{aligned}$$

with $a_{N-1} = (2[g]_1 E_2 + 2C_g C_{t^*}[\lambda]_1 (2 + C_{t^*} C_\lambda) + (4C_g C_\lambda [t^*] + 2[g]_* [Q]) \vee (3[g]_1))$.

PROOF. As before, the strong Markov property of the process $\{X(t)\}$ yields

$$\begin{aligned} \bar{v}_{N-1}(Z_{N-1}) &= \mathbf{E}[g(X_{r_{N,\beta}(Z_{N-1}, S_{N-1})}) \mathbf{1}_{\{S_N > r_{N,\beta}(Z_{N-1}, S_{N-1})\}} | Z_{N-1}] \\ &\quad + \mathbf{E}[g(Z_N) \mathbf{1}_{\{S_N \leq r_{N,\beta}(Z_{N-1}, S_{N-1})\}} | Z_{N-1}] \\ &= \mathbf{1}_{\{r_{N,\beta}(Z_{N-1}, S_{N-1}) \geq t^*(Z_{N-1})\}} K g(Z_{N-1}) \\ &\quad + \mathbf{1}_{\{r_{N,\beta}(Z_{N-1}, S_{N-1}) < t^*(Z_{N-1})\}} J(g, g)(Z_{N-1}, r_{N,\beta}(Z_{N-1}, S_{N-1})). \end{aligned}$$

The rest of the proof is similar to that of the previous theorem. \square

As in Section 5, it is now clear that an adequate choice of discretization parameters yields arbitrarily small errors if one uses the stopping-time τ_N .

7. Example. Now we apply the procedures described in Sections 4 and 6 on a simple PDMP and present numerical results.

Set $E = [0, 1[$ and $\partial E = \{1\}$. The flow is defined on $[0, 1]$ by $\phi(x, t) = x + vt$ for some positive v , the jump rate is defined on $[0, 1]$ by $\lambda(x) = \beta x^\alpha$, with $\beta > 0$ and $\alpha \geq 1$, and for all $x \in [0, 1]$, one sets $Q(x, \cdot)$ to be the uniform law on $[0, 1/2]$. Thus the process moves with constant speed v toward 1, but the closer it gets to the boundary 1, the higher the probability to jump backward on $[0, 1/2]$. Figure 1 shows two trajectories of this process for $x_0 = 0$, $v = \alpha = 1$ and $\beta = 3$ and up to the 10th jump.

The reward function g is defined on $[0, 1]$ by $g(x) = x$. Our assumptions are clearly satisfied, and we are even in the special case when the flow is Lipschitz-continuous (see Remark A.8). All the constants involved in Theorems 5.1 and 6.3 can be computed explicitly.

The real value function $V_0 = v_0(x_0)$ is unknown, but, as our stopping rule τ_N is a stopping time dominated by T_N , one clearly has

$$(7.1) \quad \begin{aligned} \bar{V}_0 = \mathbf{E}_{x_0}[g(X(\tau_N))] &\leq V_0 = \sup_{\tau \in \mathcal{M}_N} \mathbf{E}_{x_0}[g(X(\tau))] \\ &\leq \mathbf{E}_{x_0} \left[\sup_{0 \leq t \leq T_N} g(X(t)) \right]. \end{aligned}$$

The first and last terms can be evaluated by Monte Carlo simulations, which provide another indicator of the sharpness of our numerical procedure. For 10^6 Monte Carlo simulations, one obtains $\mathbf{E}_{x_0}[\sup_{0 \leq t \leq T_N} g(X(t))] = 0.9878$. Simulation results (for $d = 2$, $x_0 = 0$, $v = \alpha = 1$, $\beta = 3$, up to the 10th jump and for 10^5 Monte

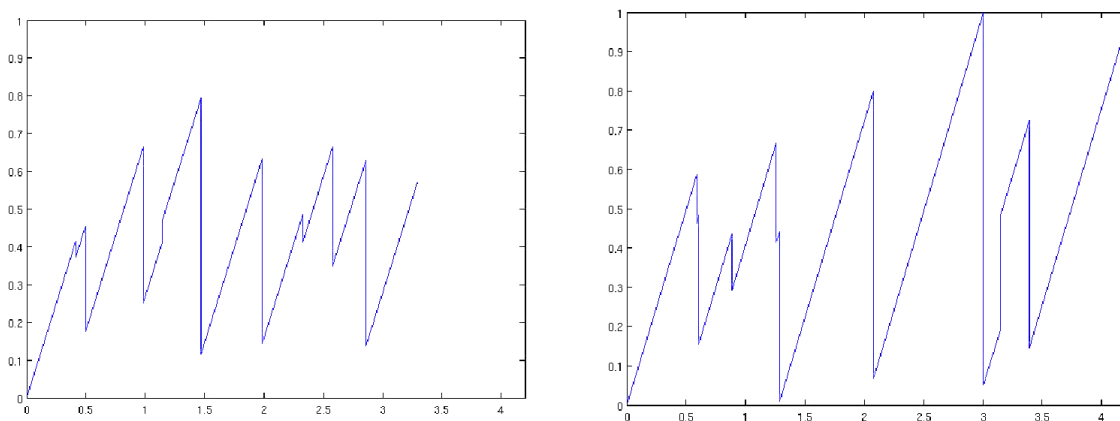


FIG. 1. Two trajectories of the PDMP.

TABLE 1
Simulation results

Pt	QE	Δ	\widehat{V}_0	\overline{V}_0	B_1	B_2	B_3
10	0.0943	0.151	0.7760	0.8173	0.1705	74.64	897.0
50	0.0418	0.100	0.8298	0.8785	0.1093	43.36	511.5
100	0.0289	0.083	0.8242	0.8850	0.1028	34.15	400.3
500	0.0133	0.056	0.8432	0.8899	0.0989	21.03	243.1
900	0.0102	0.049	0.8514	0.8968	0.0910	17.98	206.9

Pt	Number of points in each quantization grid
QE	Quantization error: $QE = \max_{0 \leq k \leq N} \ \Theta_k - \widehat{\Theta}_k\ _2$
Δ	For all z , $\Delta(z) = \Delta$
B_1	Empirical bound $\mathbb{E}_{x_0}[\sup_{0 \leq t \leq T_N} g(X(t))] - \overline{V}_0$
B_2	Theoretical bound given by Theorem 5.1
B_3	Theoretical bound given by Theorems 6.3 and 6.4

Carlo simulations) are given in Table 1. Note that, as expected, the theoretical errors decrease as the quantization error decreases. From (7.1), it follows that

$$V_0 - \overline{V}_0 \leq \mathbb{E}_{x_0} \left[\sup_{0 \leq t \leq T_N} g(X(t)) \right] - \overline{V}_0.$$

This provides an empirical upper bound for the error.

APPENDIX A: AUXILIARY RESULTS

A.1. Lipschitz properties of J and K . In this section, we derive useful Lipschitz-type properties of operators J and K . The first result is straightforward.

LEMMA A.1. *Let $h \in \mathbf{L}^c$. Then for all $(x, y) \in E^2$ and $(t, u) \in \mathbb{R}_+^2$, one has*

$$\begin{aligned} & |h(\phi(x, t \wedge t^*(x)))e^{-\Lambda(x, t \wedge t^*(x))} - h(\phi(y, u \wedge t^*(y)))e^{-\Lambda(y, u \wedge t^*(y))}| \\ & \leq D_1(h)|x - y| + D_2(h)|t - u|, \end{aligned}$$

where:

- if $t < t^*(x)$ and $u < t^*(y)$,

$$D_1(h) = [h]_1 + C_h C_{t^*}[\lambda]_1, \quad D_2(h) = [h]_2 + C_h C_\lambda,$$

- if $t = t^*(x)$ and $u = t^*(y)$,

$$D_1(h) = [h]_* + C_h C_{t^*}[\lambda]_1 + C_h C_\lambda[t^*], \quad D_2(h) = 0,$$

- otherwise,

$$D_1(h) = [h]_1 + C_h C_{t^*}[\lambda]_1 + [h]_2[t^*] + C_h C_\lambda[t^*], \quad D_2(h) = [h]_2 + C_h C_\lambda.$$

LEMMA A.2. *Let $w \in \mathbf{B}(E)$. Then for all $x \in E$, $(t, u) \in \mathbb{R}_+^2$, one has*

$$|J(w, g)(x, t) - J(w, g)(x, u)| \leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) |t - u|.$$

PROOF. By definition of J , we obtain

$$\begin{aligned} & |J(w, g)(x, t) - J(w, g)(x, u)| \\ & \leq \left| \int_{t \wedge t^*(x)}^{u \wedge t^*(x)} \lambda Qw(\phi(x, s)) e^{-\Lambda(x, s)} ds \right| \\ & \quad + |g(\phi(x, t \wedge t^*(x))) e^{-\Lambda(x, t \wedge t^*(x))} - g(\phi(x, u \wedge t^*(x))) e^{-\Lambda(x, u \wedge t^*(x))}|. \end{aligned}$$

Applying Lemma A.1 to $h = g$, the result follows. \square

LEMMA A.3. *Let $w \in \mathbf{L}^c$. Then for all $(x, y) \in E^2$, $t \in \mathbb{R}_+$,*

$$|J(w, g)(x, t) - J(w, g)(y, t)| \leq (C_w E_1 + [w]_1 E_2 + E_3) |x - y|,$$

where

$$\begin{aligned} E_1 &= C_\lambda [t^*] + C_{t^*} [\lambda]_1 (1 + C_{t^*} C_\lambda), \\ E_2 &= C_{t^*} C_\lambda [Q], \\ E_3 &= [g]_1 + [g]_2 [t^*] + C_g \{C_{t^*} [\lambda]_1 + C_\lambda [t^*]\}. \end{aligned}$$

PROOF. Again by definition, we obtain

$$\begin{aligned} & |J(w, g)(x, t) - J(w, g)(y, t)| \\ & \leq \left| \int_0^{t \wedge t^*(x)} \lambda Qw(\phi(x, s)) e^{-\Lambda(x, s)} ds - \int_0^{t \wedge t^*(y)} \lambda Qw(\phi(y, s)) e^{-\Lambda(y, s)} ds \right| \\ & \quad + |g(\phi(x, t \wedge t^*(x))) e^{-\Lambda(x, t \wedge t^*(x))} - g(\phi(y, t \wedge t^*(y))) e^{-\Lambda(y, t \wedge t^*(y))}|. \end{aligned}$$

Without loss of generality it can be assumed that $t^*(x) \leq t^*(y)$. From Lemma A.1 for $h = g$ and using the fact that $|t \wedge t^*(x) - t \wedge t^*(y)| \leq |t^*(x) - t^*(y)|$, we get

$$\begin{aligned} & |J(w, g)(x, t) - J(w, g)(y, t)| \\ & \leq \int_0^{t \wedge t^*(x)} |\lambda Qw(\phi(x, s)) e^{-\Lambda(x, s)} - \lambda Qw(\phi(y, s)) e^{-\Lambda(y, s)}| ds \\ & \quad + (C_w C_\lambda [t^*] + E_3) |x - y|. \end{aligned}$$

By using a similar results as Lemma A.1 for $h = \lambda Qw$, we obtain the result. \square

LEMMA A.4. *Let $w \in \mathbf{L}^c$. Then for all $(x, y) \in E^2$,*

$$|Kw(x) - Kw(y)| \leq (C_w E_4 + [w]_1 E_2 + [w]_* [Q]) |x - y|,$$

where $E_4 = 2C_\lambda [t^*] + C_{t^*} [\lambda]_1 (2 + C_{t^*} C_\lambda)$.

PROOF. The proof is similar to the previous ones and is therefore omitted. \square

A.2. Lipschitz properties of the value functions. Now we turn to the Lipschitz continuity of the sequence of value functions (v_n) . Namely, we prove that under our assumptions, v_n belongs to \mathbf{L}^c for all $0 \leq n \leq N$. We also compute the Lipschitz constant of v_n on \bar{E} as it is much sharper in this case than $[v_n]_1$ (see Remark 2.2).

We start with proving sharper results on operator J .

LEMMA A.5. *Let $w \in \mathbf{L}^c$. Then for all $x \in E$ and $(s, t) \in \mathbb{R}_+^2$,*

$$\left| \sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u) \right| \leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) |t - s|.$$

PROOF. Without loss of generality it can be assumed that $t \leq s$. Therefore, one has

$$\begin{aligned} \text{(A.1)} \quad & \left| \sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u) \right| \\ &= \sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u). \end{aligned}$$

Note that there exists $\bar{t} \in [t \wedge t^*(x), t^*(x)]$ such that $\sup_{u \geq t} J(w, g)(x, u) = J(w, g)(x, \bar{t})$. Consequently, if $\bar{t} \geq s$ then one has $|\sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u)| = 0$.

Now if $\bar{t} \in [t \wedge t^*(x), s[$, then one has

$$\sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u) \leq J(w, g)(x, \bar{t}) - J(w, g)(x, s).$$

From Lemma A.2, we obtain the following inequality:

$$\text{(A.2)} \quad \sup_{u \geq t} J(w, g)(x, u) - \sup_{u \geq s} J(w, g)(x, u) \leq (C_w C_\lambda + [g]_2 + C_g C_\lambda) |\bar{t} - s|.$$

Combining (A.1), (A.2) and the fact that $|\bar{t} - s| \leq |t - s|$ the result follows. \square

Similarly, we obtain the following result.

LEMMA A.6. *Let $w \in \mathbf{L}^c$. Then for all $(x, y) \in E^2$,*

$$\left| \sup_{t \leq t^*(x)} J(w, g)(x, t) - \sup_{t \leq t^*(y)} J(w, g)(y, t) \right| \leq (C_w E_5 + [w]_1 E_2 + E_6) |x - y|,$$

where $E_5 = E_1 + C_\lambda [t^*]$ and $E_6 = E_3 + ([g]_2 + C_g C_\lambda) [t^*]$.

Now we turn to (v_n) . Recall from [11] that for all $0 \leq n \leq N$, (v_n) is bounded with $C_{v_n} = C_g$.

PROPOSITION A.7. For all $0 \leq n \leq N$, $v_n \in \mathbf{L}^c$ and

$$(A.3) \quad [v_n]_1 \leq e^{C_\lambda C_{t^*}} (2[v_{n+1}]_1 E_2 + C_g E_1 + C_g E_4 + C_g C_{t^*} [\lambda]_1 (1 + C_\lambda C_{t^*})) + e^{C_\lambda C_{t^*}} \{([g]_1 + [g]_2 [t^*]) \vee ([v_{n+1}]_* [Q])\},$$

$$(A.4) \quad [v_n]_2 \leq e^{C_\lambda C_{t^*}} \{C_g C_\lambda (4 + C_\lambda C_{t^*}) + [g]_2\},$$

$$[v_n]_* \leq [v_n]_1 + [v_n]_2 [t^*],$$

$$[v_n] \leq [v_{n+1}]_1 E_2 + C_g E_5 + \{E_6 \vee ([v_{n+1}]_* [Q] + C_g C_{t^*} [\lambda]_1)\}.$$

PROOF. Clearly, $v_N = g$ is in \mathbf{L}^c . Assume that v_{n+1} is in \mathbf{L}^c , then by using the semi-group property of the drift ϕ it can be shown that for any $x \in E$, $t \in [0, t^*(x)]$, one has (see [11], equation (8))

$$(A.5) \quad v_n(\phi(x, t)) = e^{\Lambda(x, t)} \left\{ \left(\sup_{u \geq t} J(v_{n+1}, g)(x, u) \vee K v_{n+1}(x) \right) - I v_{n+1}(x, t) \right\}.$$

Note that for $x \in E$, $t \in \mathbb{R}_+$, one has

$$(A.6) \quad \begin{aligned} & \sup_{u \geq t} J(v_{n+1}, g)(x, u) \vee K v_{n+1}(x) \\ & \leq \sup_u J(v_{n+1}, g)(x, u) \vee K v_{n+1}(x) \\ & = v_n(x). \end{aligned}$$

Set $(x, y) \in E^2$ and $t \in [0, t^*(x) \wedge t^*(y)]$. It is easy to show that

$$(A.7) \quad |e^{\Lambda(x, t)} - e^{\Lambda(y, t)}| \leq e^{C_\lambda C_{t^*}} [\lambda]_1 C_{t^*} |x - y|,$$

$$(A.8) \quad |I v_{n+1}(x, t) - I v_{n+1}(y, t)| \leq (C_{v_{n+1}} E_1 + [v_{n+1}]_1 E_2) |x - y|.$$

Then, (A.5)–(A.8) yield

$$(A.9) \quad \begin{aligned} & |v_n(\phi(x, t)) - v_n(\phi(y, t))| \\ & \leq \{|v_n(x)| + |I v_{n+1}(x, t)|\} e^{C_\lambda C_{t^*}} [\lambda]_1 C_{t^*} |x - y| \\ & \quad + e^{\Lambda(y, t)} \left\{ \sup_{u \geq t} |J(v_{n+1}, g)(x, u) - J(v_{n+1}, g)(y, u)| \right. \\ & \quad \left. \vee |K v_{n+1}(x) - K v_{n+1}(y)| \right\} \\ & \quad + e^{\Lambda(y, t)} (C_{v_{n+1}} E_1 + [v_{n+1}]_1 E_2) |x - y|. \end{aligned}$$

For $x \in E$, $t \in [0, t^*(x)]$ and $n \in \mathbb{N}$, note that

$$(A.10) \quad \begin{aligned} & e^{\Lambda(x, t)} \leq e^{C_\lambda C_{t^*}}, \\ & |I v_{n+1}(x, t)| \leq C_\lambda C_{v_{n+1}} C_{t^*} \quad \text{and} \quad |v_{n+1}(x)| \leq C_g. \end{aligned}$$

Therefore, we obtain inequality (A.3) by using (A.9), (A.10) and Lemma A.3, A.5, and the fact that $C_g E_1 + E_3 = C_g E_4 + [g]_1 + [g]_2[t^*]$.

Now, set $x \in E$ and $t, s \in [0, t^*(x)]$. Similarly, one has

$$(A.11) \quad |e^{\Lambda(x,t)} - e^{\Lambda(x,s)}| \leq e^{C_\lambda C_{t^*}} C_\lambda |t - s|,$$

$$(A.12) \quad |Iv_{n+1}(x, t) - Iv_{n+1}(x, s)| \leq C_\lambda C_{v_{n+1}} |t - s|.$$

Combining (A.5), (A.6), (A.11) and (A.12), it yields

$$(A.13) \quad \begin{aligned} & |v_n(\phi(x, t)) - v_n(\phi(x, s))| \\ & \leq \{|v_n(x)| + |Iv_{n+1}(x, t)|\} e^{C_\lambda C_{t^*}} C_\lambda |t - s| \\ & \quad + e^{\Lambda(x,t)} \left\{ \left| \sup_{u \geq t} J(v_{n+1}, g)(x, u) - \sup_{u \geq s} J(v_{n+1}, g)(x, u) \right| \right. \\ & \quad \left. + C_\lambda C_{v_{n+1}} |t - s| \right\}. \end{aligned}$$

Finally, inequality (A.4) follows from equations (A.10), (A.13) and Lemma A.4.

One clearly has $[v_n]_* \leq [v_n]_1 + [v_n]_2[t^*]$. Finally, set $(x, y) \in \bar{E}^2$. By definition, one has

$$\begin{aligned} & |v_n(x) - v_n(y)| \\ & \leq \left| \sup_{u \leq t^*(x)} J(v_{n+1}, g)(x, u) - \sup_{u \leq t^*(y)} J(v_{n+1}, g)(y, u) \right| \\ & \quad \vee |Kv_{n+1}(x) - Kv_{n+1}(y)| \end{aligned}$$

and we conclude using Lemmas A.6 and A.4, and the fact that $E_4 = E_5 + C_{t^*}[\lambda]_1$. \square

REMARK A.8. Note that $[v_n]$ is much sharper than $[v_n]_1$. If in addition to our assumptions, the drift ϕ is Lipschitz-continuous in both variables, then with obvious notation, one has $[v_n]_i \leq [v_n][\phi]_i$ for $i \in \{1, 2, *\}$, which should yield better constants (see, e.g., Section 7).

APPENDIX B: STRUCTURE OF THE STOPPING TIMES OF PDMPs

Let τ be an $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ -stopping time. Let us recall the important result from Davis [6].

THEOREM B.1. *There exists a sequence of nonnegative random variables $(R_n)_{n \in \mathbb{N}^*}$ such that R_n is $\mathcal{F}_{T_{n-1}}$ -measurable and $\tau \wedge T_{n+1} = (T_n + R_{n+1}) \wedge T_{n+1}$ on $\{\tau \geq T_n\}$.*

LEMMA B.2. Define $\bar{R}_1 = R_1$, and $\bar{R}_k = R_k \mathbf{1}_{\{S_{k-1} \leq \bar{R}_{k-1}\}}$. Then one has

$$\tau = \sum_{n=1}^{\infty} \bar{R}_n \wedge S_n.$$

PROOF. Clearly, on $\{T_k \leq \tau < T_{k+1}\}$, one has $R_j \geq S_j$ and $R_{k+1} < S_{k+1}$ for all $j \leq k$. Consequently, by definition $\bar{R}_j = R_j$ for all $j \leq k + 1$, whence

$$\begin{aligned} \sum_{n=1}^{\infty} \bar{R}_n \wedge S_n &= \sum_{n=1}^k \bar{R}_n \wedge S_n + \{\bar{R}_{k+1} \wedge S_{k+1}\} + \sum_{n=k+2}^{\infty} \bar{R}_n \wedge S_n \\ &= T_k + R_{k+1} + \sum_{n=k+2}^{\infty} \bar{R}_n \wedge S_n. \end{aligned}$$

Since $\bar{R}_{k+1} = R_{k+1} < S_{k+1}$ we have $\bar{R}_j = 0$ for all $j \geq k + 2$. Therefore, $\sum_{n=1}^{\infty} \bar{R}_n \wedge S_n = T_k + R_{k+1} = \tau$, showing the result. \square

There exists a sequence of measurable mappings $(r_k)_{k \in \mathbb{N}_*}$ defined on $E \times (\mathbb{R}_+ \times E)^{k-1}$ with value in \mathbb{R}_+ satisfying

$$R_1 = r_1(Z_0),$$

$$R_k = r_k(Z_0, \Gamma_{k-1}),$$

where $\Gamma_k = (S_1, Z_1, \dots, S_k, Z_k)$.

DEFINITION B.3. Consider $p \in \mathbb{N}_*$. Let $(\hat{R}_k)_{k \in \mathbb{N}_*}$ be a sequence of mappings defined on $E \times (\mathbb{R}_+ \times E)^p \times \Omega$ with value in \mathbb{R}_+ defined by

$$\hat{R}_1(y, \gamma, \omega) = r_{p+1}(y, \gamma)$$

and for $k \geq 2$

$$\hat{R}_k(y, \gamma, \omega) = r_{p+k}(y, \gamma, \Gamma_{k-1}(\omega)) \mathbf{1}_{\{S_{k-1} \leq \hat{R}_{k-1}\}}(y, \gamma, \omega).$$

PROPOSITION B.4. Assume that $T_p \leq \tau \leq T_N$. Then, one has

$$\tau = T_p + \hat{\tau}(Z_0, \Gamma_p, \theta_{T_p}),$$

where $\hat{\tau}: E \times (\mathbb{R}_+ \times E)^p \times \Omega \rightarrow \mathbb{R}_+$ is defined by

$$(B.1) \quad \hat{\tau}(y, \gamma, \omega) = \sum_{n=1}^{N-p} \hat{R}_n(y, \gamma, \omega) \wedge S_n(\omega).$$

PROOF. First, let us prove by induction that for $k \in \mathbb{N}_*$, one has

$$(B.2) \quad \widehat{R}_k(Z_0, \Gamma_p, \theta_{T_p}) = \overline{R}_{p+k}.$$

Indeed, one has $\widehat{R}_1(Z_0, \Gamma_p, \theta_{T_p}) = R_{p+1}$, and on the set $\{\tau \geq T_p\}$, one also has $R_{p+1} = \overline{R}_{p+1}$. Consequently, $\widehat{R}_1(Z_0, \Gamma_p) = \overline{R}_{p+1}$. Now assume that $\widehat{R}_k(Z_0, \Gamma_p, \theta_{T_p}) = \overline{R}_{p+k}$. Then, one has

$$\begin{aligned} & \widehat{R}_{k+1}(Z_0(\omega), \Gamma_p(\omega), \theta_{T_p}(\omega)) \\ &= r_{p+k+1}(Z_0(\omega), \Gamma_p(\omega), \Gamma_k(\theta_{T_p}(\omega))) \mathbf{1}_{\{S_k \leq \widehat{R}_k\}}(Z_0(\omega), \Gamma_p(\omega), \theta_{T_p}(\omega)). \end{aligned}$$

By definition, one has $\Gamma_k(\theta_{T_p}(\omega)) = (S_{p+1}(\omega), Z_{p+1}(\omega), \dots, S_{p+k}(\omega), Z_{p+k}(\omega))$ and the induction hypothesis easily yields $\mathbf{1}_{\{S_k \leq \widehat{R}_k\}}(Z_0(\omega), \Gamma_p(\omega), \theta_{T_p}(\omega)) = \mathbf{1}_{\{S_{p+k} \leq \overline{R}_{p+k}\}}(\omega)$. Therefore, we get $\widehat{R}_{k+1}(Z_0, \Gamma_p, \theta_{T_p}) = \overline{R}_{p+k+1}$, showing (B.2).

Combining (B.1) and (B.2) yields

$$(B.3) \quad \widehat{\tau}(Z_0, \Gamma_p, \theta_{T_p}) = \sum_{n=1}^{N-n} \overline{R}_{p+n} \wedge S_{p+n}.$$

However, we have already seen that on the set $\{T \geq T_p\}$, one has $R_k = \overline{R}_k \geq S_k$, for $k \leq p$. Consequently, using (B.3), we obtain

$$T_p + \widehat{\tau}(Z_0, \Gamma_p, \theta_{T_p}) = \sum_{k=1}^p S_k + \sum_{k=p+1}^N \overline{R}_k \wedge S_k = \sum_{k=1}^N \overline{R}_k \wedge S_k.$$

Since $\tau \leq T_N$, we obtain from Lemma B.2 and its proof that $\tau = \sum_{n=1}^N \overline{R}_n \wedge S_n$, showing the result. \square

PROPOSITION B.5. Let $(U_n)_{n \in \mathbb{N}^*}$ be a sequence of nonnegative random variables such that U_n is $\mathcal{F}_{T_{n-1}}$ -measurable and $U_{n+1} = 0$ on $\{S_n > U_n\}$, for all $n \in \mathbb{N}_*$. Set

$$U = \sum_{n=1}^{\infty} U_n \wedge S_n.$$

Then U is an $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ -stopping time.

PROOF. Assumption 2.1 yields

$$(B.4) \quad \begin{aligned} \{U \leq t\} &= \bigcup_{n=0}^{\infty} [(\{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{t < T_{n+1}\}) \\ &\cup (\{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{T_{n+1} \leq t\})]. \end{aligned}$$

From the definition of U_n , one has $\{U \geq T_n\} = \{U_n \geq S_n\}$; hence one has

$$\begin{aligned} & \{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{t < T_{n+1}\} \\ &= \{S_n \leq U_n\} \cap \{T_n + U_{n+1} \leq t\} \cap \{T_n \leq t\} \cap \{t < T_{n+1}\}. \end{aligned}$$

Theorem 2.10(ii) in [8] now yields $\{S_n \leq U_n\} \cap \{T_n + U_{n+1} \leq t\} \cap \{T_n \leq t\} \in \mathcal{F}_t$; thus one has

$$(B.5) \quad \{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{t < T_{n+1}\} \in \mathcal{F}_t.$$

On the other hand, one has

$$\begin{aligned} & \{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{T_{n+1} \leq t\} \\ &= \{S_n \leq U_n\} \cap \{U_{n+1} < S_{n+1}\} \cap \{T_{n+1} \leq t\}. \end{aligned}$$

Hence Theorem 2.10(ii) in [8] again yields

$$(B.6) \quad \{T_n \leq U < T_{n+1}\} \cap \{U \leq t\} \cap \{T_{n+1} \leq t\} \in \mathcal{F}_t.$$

Combining equations (B.4), (B.5) and (B.6) we obtain the result. \square

COROLLARY B.6. *For any $(y, \gamma) \in E \times (\mathbb{R}_+ \times E)^p$, $\widehat{\tau}(y, \gamma, \cdot)$ is an $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ -stopping time satisfying $\widehat{\tau}(y, \gamma, \cdot) \leq T_{N-p}$.*

PROOF. It follows from the definition of \widehat{R}_k that $\widehat{R}_k(y, \gamma, \omega) < S_k(\omega)$ implies $\widehat{R}_{k+1}(y, \gamma, \omega) = 0$ and the nonnegative random variable $\widehat{R}_k(y, \gamma, \cdot)$ is $\mathcal{F}_{T_{k-1}}$ -measurable. Therefore, Proposition B.5 yields that $\widehat{\tau}(y, \gamma, \cdot)$ is an $\{\mathcal{F}_t\}_{t \in \mathbb{R}_+}$ -stopping time. Finally, by definition of $\widehat{\tau}$ [see (B.1)], one has $\widehat{\tau}(y, \gamma, \cdot) \leq \sum_{n=1}^{N-p} S_n = T_{N-p}$ showing the result. \square

REFERENCES

- [1] BALLY, V. and PAGÈS, G. (2003). A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli* **9** 1003–1049. [MR2046816](#)
- [2] BALLY, V., PAGÈS, G. and PRINTEMS, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance* **15** 119–168. [MR2116799](#)
- [3] COSTA, O. L. V. and DAVIS, M. H. A. (1988). Approximations for optimal stopping of a piecewise-deterministic process. *Math. Control Signals Systems* **1** 123–146. [MR936330](#)
- [4] COSTA, O. L. V. and DUFOUR, F. (2008). Stability and ergodicity of piecewise deterministic Markov processes. *SIAM J. Control Optim.* **47** 1053–1077. [MR2385873](#)
- [5] COSTA, O. L. V., RAYMUNDO, C. A. B. and DUFOUR, F. (2000). Optimal stopping with continuous control of piecewise deterministic Markov processes. *Stochastics Stochastics Rep.* **70** 41–73. [MR1785064](#)
- [6] DAVIS, M. H. A. (1993). *Markov Models and Optimization. Monographs on Statistics and Applied Probability* **49**. Chapman and Hall, London. [MR1283589](#)
- [7] DUFOUR, F. and COSTA, O. L. V. (1999). Stability of piecewise-deterministic Markov processes. *SIAM J. Control Optim.* **37** 1483–1502. [MR1710229](#)
- [8] ELLIOTT, R. J. (1982). *Stochastic Calculus and Applications. Applications of Mathematics (New York)* **18**. Springer, New York. [MR678919](#)

- [9] GAȚAREK, D. (1991). On first-order quasi-variational inequalities with integral terms. *Appl. Math. Optim.* **24** 85–98. [MR1106927](#)
- [10] GRAY, R. M. and NEUHOFF, D. L. (1998). Quantization. *IEEE Trans. Inform. Theory* **44** 2325–2383. [MR1658787](#)
- [11] GUGERLI, U. S. (1986). Optimal stopping of a piecewise-deterministic Markov process. *Stochastics* **19** 221–236. [MR872462](#)
- [12] KUSHNER, H. J. (1977). *Probability Methods for Approximations in Stochastic Control and for Elliptic Equations. Mathematics in Science and Engineering* **129**. Academic Press, New York. [MR0469468](#)
- [13] LENHART, S. and LIAO, Y. C. (1985). Integro-differential equations associated with optimal stopping time of a piecewise-deterministic process. *Stochastics* **15** 183–207. [MR869199](#)
- [14] PAGÈS, G. (1998). A space quantization method for numerical integration. *J. Comput. Appl. Math.* **89** 1–38. [MR1625987](#)
- [15] PAGÈS, G. and PHAM, H. (2005). Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli* **11** 893–932. [MR2172846](#)
- [16] PAGÈS, G., PHAM, H. and PRINTEMS, J. (2004). An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems. *Stoch. Dyn.* **4** 501–545. [MR2102752](#)
- [17] PAGÈS, G., PHAM, H. and PRINTEMS, J. (2004). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of Computational and Numerical Methods in Finance* 253–297. Birkhäuser, Boston, MA. [MR2083055](#)

B. DE SAPORTA
F. DUFOUR
K. GONZALEZ
UNIVERSITÉ DE BORDEAUX
IMB, 351 COURS DE LA LIBÉRATION
F33405 TALENCE CEDEX
FRANCE
E-MAIL: dufour@math.u-bordeaux1.fr
saporta@math.u-bordeaux1.fr
gonzalez@math.u-bordeaux1.fr

- 2.2 [BBGP⁺13] C. BAYSSE, D. BIHANNIC, A. GÉGOUT PETIT, M. PRENAT, B. DE SAPORTA, AND J. SARACCO. Maintenance optimization of optronic equipment. In *Prognostics and System Health Management Conference*, Milano, Italy, 2013.

Maintenance optimisation of optronic equipment

C. Baysse^{a,b}, D. Bihannic^b, A. Gegout-Petit^a, M. Prenat^b, B. de Saporta^a, J. Saracco^a

^aThales Optronique, 2 Avenue Gay Lussac, 78990 Élancourt France.

^bBordeaux University, UMR CNRS 5251 IMB, INRIA Bordeaux Sud-ouest CQFD, 200, avenue de la Vieille Tour 33405 Talence cedex France.

camille.baysse@inria.fr

As part of optimizing the reliability, Thales Optronics now includes systems that examine the state of its equipment. This function is performed by HUMS (Health & Usage Monitoring System). The aim is to implement in the HUMS a program based on observations that can determine the state of the system and propose a maintenance action before failures. So we decompose our problem into two steps: the first step is to detect the degraded state (which announces future failure) using an informative variable and hidden Markov chains. This step was developed in Baysse & al (2012). The second is to propose an optimal and dynamic maintenance policy, adapted to the state of the system and taking into account both random failures and those related to the degradation phenomenon. We want to estimate the best time to perform maintenance: a maintenance performed too early may be unnecessarily costly and inconvenient for the client but too late may cause the occurrence of a failure that will damage the rest of the equipment and may be responsible for the failure of a mission. So it is necessary to find a balance between these two extreme maintenance policies. First, we model the state of the system by a piecewise-deterministic Markov process: PDMP (introduced by Davis 1993). Often the evolution of the system is modeled by stochastic processes such as Markov jump process, semi-Markov process (Cocozza & al (1997)). There are also tools for modeling such as Stochastic Petri networks (Marsan & al 1995), dynamic Bayesian networks (Donnat & al 2010). However, the flexibility of modeling by PDMP allows to take into account the dynamic component degradation. The works of Lair & al (2012) focuses on this topic, they use a finite volume scheme to evaluate the quantities of interest associated with PDMP. Even if there are different methods that optimize maintenance policy, few use optimal stopping. In this paper, we use this method whose principle is to maximize a performance function that takes into account operating time, maintenance costs, repairs and downtime. We use the numerical probability tools developed in de Saporta & al (2012) in order to compute this conditioned-based time of maintenance. The integration of this method in the HUMS, will be soon implemented in specific optronic equipment by Thales. We present results of simulation in this case. The methodology can be extended to more complicated cases.

1. Industrial context

Thanks to the HUMS, each of the appliances has a logbook which provides information at each start-up such as: number of uses, cumulative operating time of appliance, "cool down time" (Tmf)... This Tmf is the transit time for the system from ambient temperature to a very low one. This temperature decrease is required to operate appliance and this is done on every boot. According to experts, a Tmf increase results from deterioration in the cooling system. According to this hypothesis, a careful observation of Tmf evolution allows us to determine the state of the cooling system. We suppose that the cooling system pass from stable state to degraded state and from degraded state to reach failure. In Baysse & al (2012), we have given a mathematical method based on Hidden Markov Chain in order to detect a transition to a degraded state of the cooling system. There are two other kinds of possible failures: electronic failure and ball bearing failure. These two failures do not pass by a degraded state. So we study appliance with three failures (electronic, ball bearing and cooling system failure) and three states (stable, degraded on account of cooling system, failure).

Our objective is to propose an optimal and dynamic maintenance policy adapted to the random state of the system.

2. Modeling

In order to develop a maintenance policy that takes into account both random failures and those related to a degradation phenomenon, we model the state of the system by a piecewise-deterministic Markov process. This modeling makes possible the transition from the stable state to failure directly (random case) or through the degraded state (damage to the cooling system (see figure 1).

The notion of PDMP was introduced by Davis (1993). PDMP are processes with deterministic evolution, punctuated by random jumps and changes of regimes that can allow them to pass from one state to another. PDMP are hybrid process generally noted $\xi_t = (m_t, s_t)_{t \in \mathbb{R}_+}$. The first component m_t is a discrete variable with values in a finite or countable space M . It describes the state of the system at time t (system in stable mode, degraded, failure ...). The second component s_t evolves in a continuous way in $E_m \in \mathbb{R}^n$ and describes evolution of the system in the mode m_t by its physical variables (for example pressure, age of system..).

Our study is about equipment with three states: stable state ($m_t = 1$), degraded state ($m_t = 2$) and failure ($m_t = 3$).

At the beginning equipment is in stable state and then it breakdowns or it goes in degraded state:

- If it breakdowns directly, it is due to an electronic failure or a failure about ball bearing. Failure rates are respectively λ_1 and $\lambda_3(t)$.
- If it goes from stable state to a degraded state, it is due to a deterioration of the cooling system. The occurrence rate of this deterioration is noted λ_0 . In the degraded state, it is possible to have electronic, ball bearing or cooling system failures (rate λ_2).

Equipment in degraded state or in failure cannot return in stable state. Note that the state "failure" is absorbant and so the number of jump is less than or equal to 2. Figure 1 illustrates how the system works. According to experts, the rate λ_3 depends on the age t of equipment contrary to other rates. Note that m_t is not markovian because rate λ_3 depends on t . Thanks to our method to detect transition from state 1 to state 2, (see Baysse & al (2012)), we suppose that the jump of the process is observed.

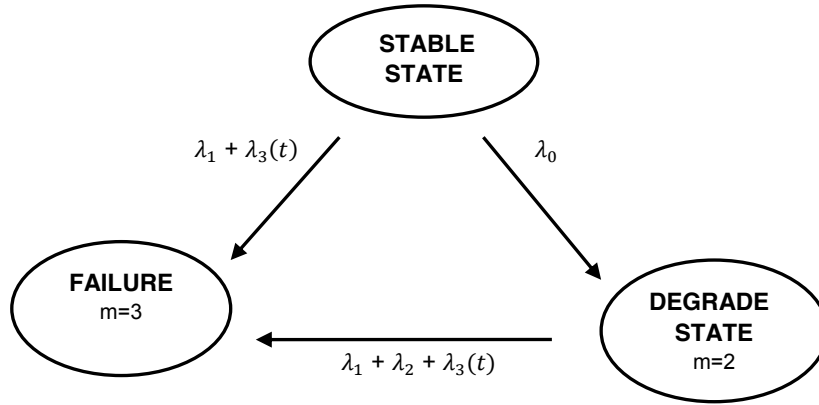


Figure 1 : Modeling system

In order to use the powerful framework of Markov process, we must add time t to the process $(m_t)_{t \in \mathbb{R}_+}$, as information such that $\xi_t = (m_t, s_t = t)_{t \in \mathbb{R}_+}$ is markovian. Indeed it is a PDMP. So the PDMP considered here describes the state of equipment and its age: $\xi_t = (m_t, s_t = t)_t$. Its motion is described by the three characteristics (see Davis (1993)):

- the flow $\phi(m, t; s) = (m, t+s)$,
- the rate of jump $\lambda(m_t, t) = (\lambda_0 + \lambda_1 + \lambda_3(t))1_{\{m_t=1\}} + (\lambda_1 + \lambda_2 + \lambda_3(t))1_{\{m_t=2\}}$,
- the measure of transition:

$$Q(m_t, t; \{e\} \times \{t\}) = \begin{cases} \left(\frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_3(t)} \right) 1_{\{e=2\}} + \left(\frac{\lambda_1 + \lambda_3(t)}{\lambda_0 + \lambda_1 + \lambda_3(t)} \right) 1_{\{e=3\}} & \text{if } m_t = 1 \\ 1_{\{e=3\}} & \text{if } m_t = 2 \end{cases}$$

We denote $Z_0 = (1, 0)$, $T_0 = 0$. The state of the system just after the first transition is $Z_1 = (m_{T_1}, T_1)$ and $Z_2 = (m_{T_2}, T_2) = (3, T_2)$ if T_2 occurs. We put $S_1 = T_1$ and $S_2 = T_2 - T_1$ the interjumping times. With these

notations, the discrete process $\theta_n = (Z_n, S_n) = (m_{T_n}, T_n, S_n)$ associated with the PDMP $(\xi_t)_t$ is a Markov chain and it is considered for $n = \{0, 1, 2\}$.

The horizon T of the study is finite. So that the remaining time is $t^*(\xi_t) = T - t$.

3. Optimal and dynamic maintenance policy

We consider the problem as an optimal stopping problem for PDMP, whose principle is to maximize a performance function that takes into account operating time, maintenance costs, repairs and downtime. We want to estimate the best time to perform maintenance in order to allow Thales to manage upstream park equipment. Recent work has been done on this subject. De Saporta & al (2012) give the theoretical foundations of the method that we use in this study. In de Saporta & al (2010), a method of computation of best time to perform maintenance on a complex dynamic system is implemented and analyzed.

Principle of optimal stopping time

Our aim is to find a stopping time τ which maximizes expectation of a performance function at random stopping time τ that is $E_{\xi_0=(m_0,0)}[g(m_\tau, \tau)]$ where g is the function of system performance and τ a stopping time adapted to the filtration of the PDMP. This problem is typically an optimal stopping problem which consists in solving the following optimization problem:

$$v_0(m_0, 0) = \sup_{\tau \leq t^*(m_0,0)} E_{\xi_0=(m_0,0)}[g(m_\tau, \tau)]$$

$$\text{with } \begin{cases} v_2 = g \\ v_1 = L(v_2, g) \end{cases} \quad (1)$$

Function v_0 is called the value function of the problem and represents the maximum performance that can be achieved. Operator L is defined by $L(w, g)(x) = \sup_{\tau \leq t^*(x)} \{E_x[w(Z_1)1_{\{S_1 < \tau\}}] + g(\phi(x; \tau))P_x(S_1 \geq \tau)\} \vee E_x[w(Z_1)]$. It is a complex operator that depends on the characteristics of the PDMP. However, we can see that it depends on the PDMP only through the underlying Markov chain θ_n . In our case, we chose the performance function $g(m_t, t) = \begin{cases} t & \text{if } m_t = 1 \text{ or } 2 \\ 0 & \text{if } m_t = 3 \end{cases}$. Here this function favours a long time of use but is canceled if the system fails. In practice, the optimal stopping time does not necessarily exist. However, we can always find time to stop that approach the optimal performance as near as you want.

Numerical method of optimal stopping

We apply methodology developed in de Saporta & al (2010).

To approximate the ε -optimal stopping time τ we introduce a sequence of random variables $(V_n)_{n \in \{0,1,2\}}$ such as $V_n = v_n(Z_n)$. This allows to replace the recurrence (1) which covers functions by a recurrence on random variables easier to treat numerically. To approximate the values of this sequence, we proceed in two steps. First, we discretize the process on a regular time grid noted $G(\xi)$ associated with interval $[0, t^*(\xi)]$, in order to obtain a discrete time Markov chain. Thus the operator L is maximized on a finite number of points and not on a continuous time interval. This new discretized operator is noted L^d . The second step is the quantization that transforms the continuous random variables θ_n into a discrete random variable $\hat{\theta}_n$. Quantization provides a finite set of points (a grid) adapted to the law of the process and not arbitrary regular basis on the state space. Details of this method are given by Pagès & al (2003). It is based on simulations of the Markov chain (θ_n) . So we denote $\hat{\theta}_n = (\hat{Z}_n, \hat{S}_n) = (\hat{m}_{\hat{T}_n}, \hat{T}_n, \hat{S}_n)$ projection of θ_n on the quantization grid Γ_n^Z . After these two steps, the operator L is approximated by operators \hat{L}_k^d for $k \in \{1, 2\}$.

Now we can build a sequence of variables (\hat{V}_n) which approaches (V_n) . To do this, we first consider the following process:

$$\begin{cases} \hat{v}_2(z) = g(z) \text{ with } z \in \Gamma_2^Z, \\ \hat{v}_1(z) = \hat{L}_2^d(\hat{v}_2, g)(z) \text{ with } z \in \Gamma_1^Z, \\ \hat{v}_0(z) = \hat{L}_1^d(\hat{v}_1, g)(z) \text{ with } z \in \Gamma_0^Z. \end{cases} \quad (2)$$

with

$$\hat{L}_k^d(w, g)(\xi) = (\max_{t \in G(\xi)} \{E[w(\hat{Z}_k)1_{\{\hat{S}_k < t\}} | \hat{Z}_{k-1} = \xi] + g(\phi(\xi; t))P(\hat{S}_k \geq t | \hat{Z}_{k-1} = \xi)\}) \vee E[w(\hat{Z}_k) | \hat{Z}_{k-1} = \xi].$$

The approximation of V_k is performed by $\hat{V}_k = \hat{v}_k(\hat{Z}_k)$ for $k \in \{0,1,2\}$. It is shown in de Saporta & al (2010) that the error of approximation of the value function $|\hat{V}_0 - V_0|$ can be made arbitrarily small by a suitable choice of discretization parameters). A stopping time arbitrarily close to the optimal is also provided.

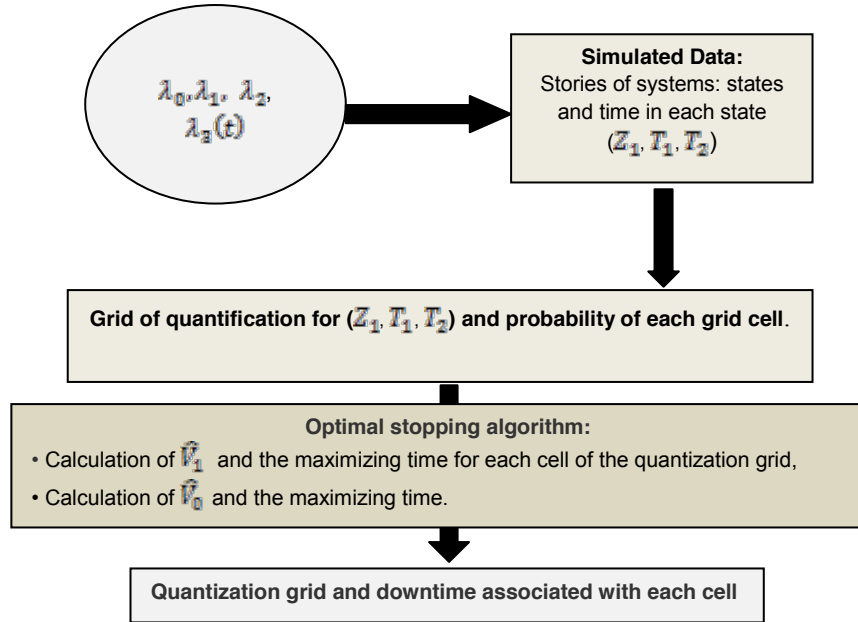


Figure 2: Schematic representation of the algorithm

4. Applications

A general presentation of the algorithm used to perform the maintenance policy is given in Figure 2. From the value of T and failure rates $\lambda_0, \lambda_1, \lambda_2, \lambda_3(t)$ provided by experts, we built a simulator of the trajectories of the process. For each of them, we have the following information: the time S_1 spent in the stable state, the type of jump ($m_{T_1}=2$ or 3) and the time spent in the new state if $m_{T_1} = 2$. From this simulation, we have created the quantization grid using an algorithm given in Pagès & al (2003). All of these elements will allow us to calculate \hat{V}_0 and \hat{V}_1 of each cell and the times τ_0, τ_1 which maximize them. Note that τ_0 is deterministic and τ_1 depends on the cell. So for each cell of the quantization grid we can associate a time nearly optimal. Let us remark that the result of this algorithm only depends on T and the failure rates and it is compute once and for all.

It will suffice to project data of equipment chosen on the new grid to propose stopping time that will be associated.

In the practice, maintenance policy is the following:

- at the beginning, a maintenance date is announced at a fixed date τ_0 for all equipment,
- if an appliance goes in degraded state to the time T_1 before the date fixed τ_0 , maintenance time is recalculated and replaced by a new time τ_1 . The time τ_1 is given by the optimal stopping time associated to the cell of \hat{T}_1 (the projection of T_1 on the grid).

To illustrate this point, we choose to look at the history of 10 appliances. In parallel we launched the algorithm to build the downtime for each device. Examples of results are presented in Table 1.

Table 1: Results of simulations

Equipment (n°)	1	2	3	4	5	6	7	8	9	10
T_1	6899	3766	6802	2238	7090	3432	4162	3800	4212	2579
m_{T_1}	2	2	3	2	3	3	2	2	2	2
T_2 if $Z_1 = (2, T_1)$	6981	3834	-	2598	-	-	4309	3885	4393	2627
Maintenance date	5160	3827	5160	2508	5160	3432	4192	3860	4242	2627

We have three possible cases:

- maintenance is at time τ_0 and before the first jump (ex n°1,3,5). We can remark that in cases 3 and 5, the first jump would have resulted to a failure.
- maintenance is between the first and second jump, when the system is in a degraded $m_{T_1} = 2$ (cases n°2,4,7,8,9). Maintenance is also before failure.
- maintenance is triggered by the failure of the system (cases n°6,10). Indeed, algorithm had planned to stop equipment n° 6 at time 5160 and n° 10 at around 2679 but these two appliances broke down before this date (that is why the stopping time equals downtime). In this case the performance achieved is zero.

In this case $\tau_0 = 5160$ and if T_1 occurs before failure, we clearly see that τ_1 depends on T_1 .

5. Results

Simulation studies allowed us to estimate the proportion of equipment in each state. We simulated 100,000 stories. In Figure 3, we let the system evolve without performing maintenance. Then we obtain the following proportions:

- 18% of equipment have ball bearing failure,
- 39% of equipment have electronic failure,
- 43% of equipment go to degraded state. These equipment have subsequently a failure of the cooling system. In this case the performance equals to zero.

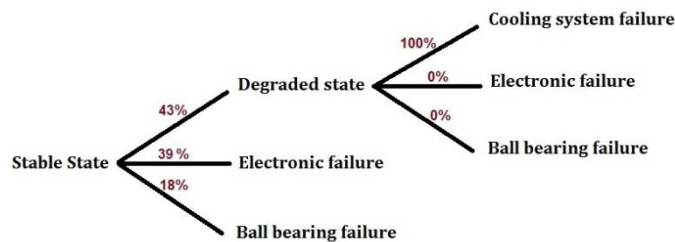


Figure 3: Evolution of the system without maintenance

Now we implement a maintenance policy. The results are given in Figure 4. At time $t = 0$, we have the first date of maintenance given by the previous algorithm. At this date, the situation is as follows :

- 30% of equipments have electronic failure before this date of maintenance,
- 8% of equipment have ball bearing failure before this date of maintenance,
- 26% of equipment go to degraded state before this date of maintenance,
- 36% of equipment are sent for maintenance at the time τ_0 . At this moment, a new date is given for maintenance equipment passed in a degraded state. For those gone in this new state 5% fail before this new date of maintenance and 95% are sent for maintenance at this new date. In this case the average performance equals to 2249.

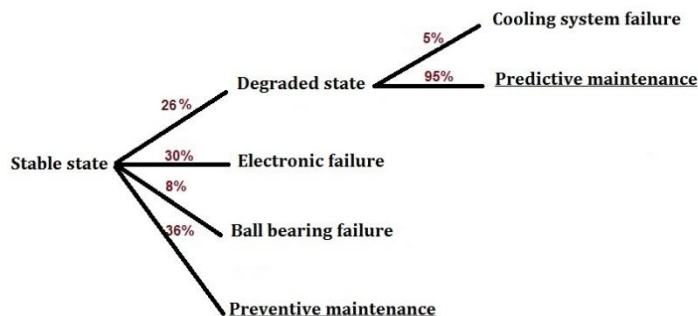


Figure 4: Evolution of the system with maintenance

So using this maintenance policy allows us to recall 61% of equipment before failure. But if we do not take into account electronic failures, 87% of equipment are sent for maintenance before failure. Indeed, we cannot perform maintenance on electronic parts, we can only replace it with a new piece. When electronic part of equipment fails, it will not damage the rest of equipment, to repair it is enough to replace as maintenance. So we do not make maintenance on electronic components.

6. Conclusion

Estimation of the state of the system associated with a decision criterion should allow to adapt maintenance policies to the observed state of the system, by the detection of failures predictable and a better management of park of equipment available. Thus, Thales will improve its maintenance action, its equipment availability at the lowest cost and the satisfaction of its customers.

References

- Baysse C., Bihannic D., Gégout-Petit A., Prenat M., Saracco J., 2012, Detection of a degraded operating mode of optronic equipment using Hidden Markov Model, ESREL, Helsinki.
- Cocozza-Thivent C., 1997, Processus stochastique et fiabilité des systèmes, Edition Springer.
- Davis M. H. A, 1993, Markov Models and Optimization, Chapman and Hall, London.
- de Saporta B., Dufour F., Zhang H., Elegbede C., 2012, Optimal stopping for the predictive maintenance of a structure subject to corrosion, Journal of Risk and Reliability, 226-(2), p169-181.
- de Saporta B., Dufour F., Gonzalez K., 2010, Numerical method for optimal stopping of piecewise deterministic Markov process, The Annals of Applied Probability, 20(5) p1607-1637.
- Donat R., Leray P., Bouillaut L., Aknin.P, 2010, A dynamic bayesian network to represent discrete duration models, Neurocomputing, 73(4-6) :570-577.
- Lair W., Mercier S., Roussignol M., Ziani.R, 2012, Dynamic modeling of complex systems for reliability calculations and maintenance optimization, Actes du congrès LM 18 - octobre 2012, Tours, France.
- Marsan A., Balbo G., Donatelli S., Franceschinis G., 1995, Modeling with generalized stochastic Petri nets, Wiley.
- Pagès G., Printems J., 2003, Optimal quadratic quantization for numerics: the Gaussian case. Monte Carlo Methods and Applications 9(2), 135-165.

- 2.3 [dSDZ⁺12] B. DE SAPORTA, F. DUFOUR, H. ZHANG, AND C. ELEGBEDE. Optimal stopping for the predictive maintenance of a structure subject to corrosion. *journal of Risk and Reliability*, 226(2) :169–181, 2012.

Optimal stopping for the predictive maintenance of a structure subject to corrosion

B de Saporta¹, F Dufour¹, H Zhang^{1*}, and C Elegbede²

¹Université de Bordeaux, INRIA Bordeaux Sud Ouest Team CQFD, Talence cedex, France

²Astrium, Saint Medard en Jalles, France

The manuscript was received on 4 February 2011 and was accepted after revision for publication on 24 May 2011.

DOI: 10.1177/1748006X11413681

Abstract: This paper presents a numerical method to compute the optimal maintenance time for a complex dynamic system applied to an example of maintenance of a metallic structure subject to corrosion. An arbitrarily early intervention may be uselessly costly, but a late one may lead to a partial/complete failure of the system, which has to be avoided. One must therefore find a balance between these too-simple maintenance policies. To achieve this aim, the system is modelled by a stochastic hybrid process. The maintenance problem thus corresponds to an optimal stopping problem. A numerical method is proposed to solve the optimal stopping problem and optimize the maintenance time for this kind of process.

Keywords: dynamic reliability, predictive maintenance, piece-wise-deterministic Markov processes, optimal stopping times, optimization of maintenance

1 INTRODUCTION

A complex system is inherently sensitive to failures of its components. Therefore maintenance policies must be determined in order to maintain an acceptable operating condition. The optimization of maintenance is a very important problem in the analysis of complex systems. It determines when maintenance tasks should be performed on the system. These intervention dates should be chosen to optimize a cost function, that is to say, maximize a performance function or, similarly, to minimize a loss function. Moreover, this optimization must take into account the random nature of failures and random evolution and dynamics of the system. Theoretical study of the optimization of maintenance is also a crucial step in the process of optimization of conception and study of the life service of the system before the first maintenance.

An example of maintenance is considered here. It is related to an aluminium metallic structure subject to corrosion. This example was provided by Astrium. It concerns a small structure within a strategic ballistic missile. The missile is stored successively in a workshop, in a nuclear submarine missile launcher in operation or in the submarine in dry-dock. These various environments are more or less corrosive and the structure is inspected with a given periodicity. It is made to have potentially large storage durations. The requirement for security is very strong. The mechanical stress exerted on the structure depends in part on its thickness. A loss of thickness will cause an over-constraint and therefore increase a risk of rupture. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the failure.

The only maintenance operation considered here is the complete replacement of the structure. Partial repairs are not allowed. Mathematically, this problem of preventive maintenance corresponds to a stochastic optimal stopping problem as explained for example in the book of Aven and Jensen [1]. It is a difficult problem, because on the one hand, the structure spends random times in each environment, and on the other hand, the

*Corresponding author: Université de Bordeaux, INRIA Bordeaux Sud-Ouest, IMB, 351 cours de la libération, 33405 Talence cedex, France.
email: huilong.zhang@math.u-bordeaux1.fr

corrosiveness of each environment is also supposed to be random within a given range. In addition, the optimal maintenance date that is searched for should be adapted to the particular history of each structure, and not an average one. The predicted maintenance date should also be updated given the past history of the corrosion process.

The recent literature on optimization of maintenance mainly focuses on comparison of policies [2–4]. A different approach is presented here. Indeed, not only a method to compute a close to optimal maintenance date is presented, but it is also proved that it becomes closer to optimal as the discretization parameters are refined.

This maintenance problem can be formulated as an optimal stopping problem for a piecewise-deterministic Markov process (PDMP). This class of problems has been studied from a theoretical point of view in reference [5]. PDMPs are a class of stochastic hybrid processes that were introduced by Davis [6] in the 1980s. These processes have two components: a Euclidean component that represents the physical system (e.g. temperature, pressure, thickness loss) and a discrete component that describes its regime of operation and/or its environment. Starting from a state x and mode m at the initial time, the process follows a deterministic trajectory given by the laws of physics until a jump time that can be either random (e.g. it corresponds to a component failure or a change of environment) or deterministic (when a magnitude reaches a certain physical threshold, for example the pressure reaches a critical value that triggers a valve). The process restarts from a new state and a new mode of operation, and so on. This defines a Markov process. Such processes can naturally take into account the dynamic and uncertain aspects of the evolution of the system. A subclass of these processes has been introduced by Devooght [7] for an application in the nuclear field. The general model has been introduced in dynamic reliability by Dufour and Dutuit [8].

The theoretical problem of optimal stopping for PDMPs is well understood, see e.g. Gugerli [9]. However, there are surprisingly few works in the literature presenting practical algorithms to compute the optimal cost and optimal stopping time. To our best knowledge only Costa and Davis [10] have presented an algorithm for calculating these quantities for PDMPs. Yet, as illustrated above, it is crucial to have an efficient numerical tool to compute the optimal maintenance time in practical cases. The objective of the present paper is to demonstrate the high practical power of the theoretical methodology described in [5]. Applying this general approach to a specific real-life industrial example

brings new technical difficulties such as the scaling problem described in section 5. More precisely, the algorithm given in this paper computes the optimal cost as well as a quasi-optimal stopping rule, that is the date when the maintenance should be performed. As a by-product of our procedure, the distribution of the optimal maintenance dates is also obtained, and dates such that the probability to perform a maintenance before this date is below a prescribed threshold can also be computed.

The remainder of this paper is organized as follows. In section 2, the example of corrosion of the metallic structure is presented with more details as well as the framework of PDMPs. In section 3, the formulation of the optimal stopping problem for PDMPs and its theoretical solution are briefly recalled. In section 4, the four main steps of the algorithm are detailed. In section 5 the numerical results obtained on the example of corrosion are presented and discussed. Finally, in section 6, a conclusion and perspectives are presented.

2 MODELLING

Throughout this paper, our approach will be illustrated on an example of maintenance of a metallic structure subject to corrosion. This example was proposed by Astrium. As explained in the introduction, it is a small homogeneous aluminium structure within a strategic ballistic missile. The missile is stored for potentially long periods in more or less corrosive environments. The mechanical stress exerted on the structure depends in part on its thickness. A loss of thickness will cause an over-constraint and therefore increase a risk of rupture. It is thus crucial to control the evolution of the thickness of the structure over time, and to intervene before the failure.

The usage profile of the missile is now described more precisely. It is stored successively in three different environments: the workshop, the submarine in operation, and the submarine in dry-dock. This is because the structure must be equipped and used in a given order. Then it goes back to the workshop and so on. The missile stays in each environment during a random duration with exponential distribution. Its parameter depends on the environment. At the beginning of its service time, the structure is treated against corrosion. The period of effectiveness of this protection is also random, with a Weibull distribution. The thickness loss only begins when this initial protection is gone. The degradation law for the thickness loss then depends on the environment through two parameters, a deterministic

transition period and a random corrosion rate uniformly distributed within a given range. Typically, the workshop and dry-dock are the more corrosive environments. The randomness of the corrosion rate accounts for small variations and uncertainties in the corrosiveness of each environment.

This degradation process is modelled by a 3D PDMP (X_t) with three modes corresponding to the three different environments. Before giving the detailed parameters of this process, general PDMPs are briefly presented.

2.1 Definition of piecewise-deterministic Markov processes

PDMPs are a general class of hybrid processes. Let M be the finite set of the possible modes of the system. In the corrosion example, the modes correspond to the various environments. For all mode m in M , let E_m be an open subset in \mathbb{R}^d . A PDMP is defined from three local characteristics (Φ, λ, Q) where

- the flow $\Phi : M \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is continuous and for all $s, t \geq 0$, one has $\Phi(\cdot, \cdot, t+s) = \Phi(\Phi(\cdot, \cdot, s), t)$. It describes the deterministic trajectory of the process between jumps. For all (m, x) in $M \times E_m$, set

$$t^*(m, x) = \inf\{t > 0 : \Phi(m, x, t) \in \partial E_m\}$$

the time to reach the boundary of the domain starting from x in mode m .

- the jump intensity λ characterizes the frequency of jumps. For all (m, x) in $M \times E_m$, and $t \leq t^*(m, x)$, set

$$\Lambda(m, x, t) = \int_0^t \lambda(\Phi(m, x, s)) ds$$

- the Markov kernel Q represents the transition measure of the process and allows one to select the new location and mode after each jump.

The trajectory $X_t = (m_t, x_t)$ of the process can then be defined iteratively. It starts at an initial point $X_0 = (k_0, y_0)$ with $k_0 \in M$ and $y_0 \in E_{k_0}$. The first jump time T_1 is determined by

$$\mathbb{P}_{(k_0, y_0)}(T_1 > t) = \begin{cases} e^{-\Lambda(k_0, y_0, t)} & \text{if } t < t^*(k_0, y_0) \\ 0 & \text{if } t \geq t^*(k_0, y_0) \end{cases}$$

On the interval $[0, T_1)$, the process follows the deterministic trajectory $m_t = k_0$ and $x_t = \Phi(k_0, y_0, t)$. At the random time T_1 , a jump occurs. Note that a

jump can be either a discontinuity in the Euclidean variable x_t or a change of mode. The process restarts at a new mode and/or position $X_{T_1} = (k_1, y_1)$, according to distribution $Q_{k_0}(\Phi(k_0, y_0, T_1), \cdot)$. An inter-jump time $T_2 - T_1$ is then selected in a similar way, and in the interval $[T_1, T_2)$ the process follows the path $m_t = k_1$ and $x_t = \Phi(k_1, y_1, t - T_1)$. Thereby, iteratively, a PDMP is constructed, see Fig. 1 for an illustration. Let $Z_0 = X_0$, and for $n \geq 1$, $Z_n = X_{T_n}$, location and mode of the process after each jump. Let $S_0 = 0$, $S_1 = T_1$ and for $n \geq 2$, $S_n = T_n - T_{n-1}$ the inter-jump times between two consecutive jumps, then (Z_n, S_n) is a Markov chain, which is the only source of randomness of the PDMP and contains all information on its random part. Indeed, if one knows the jump times and the positions after each jump, one can reconstruct the deterministic part of the trajectory between jumps. This is a very important property of PDMPs which is at the basis of the numerical procedure.

2.2 Example of corrosion of metallic structure

Now we return to the example of corrosion of structure and give the characteristics of the PDMP modelling the thickness loss. The finite set of modes is $M = \{1, 2, 3\}$, where mode 1 corresponds to the workshop environment, mode 2 to the submarine in operation, and mode 3 to the dry-dock. Although the thickness loss is a 1D process, one needs a 3D PDMP to model its evolution, because it must also take into account all the sources of randomness, that is the duration of the initial protection and the corrosion rate in each environment. The corrosion process (X_t) is defined by

$$X_t = (m_t, d_t, \gamma_t, \rho_t) \in \{1, 2, 3\} \times \mathbb{R}_+ \times \mathbb{R}_+ \times \mathbb{R}_+$$

where m_t is the environment at time t , d_t is the thickness loss at time t , γ_t is the remainder of the initial protection at time t and ρ_t is the corrosion rate of the current environment at time t .

Originally, at time 0, one has $X_0 = (1, 0, \gamma_0, \rho_0)$, which means that the missile is in the workshop and the structure has not started corroding yet. The original protection γ_0 is drawn according to a Weibull distribution function

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\beta}\right)^\alpha\right)$$

with $\alpha = 2.5$ and $\beta = 42, 40.10^6 \text{ s}^{-1}$. The corrosion rate in the workshop is drawn according to a uniform distribution on $[2, 78.10^{-7}, 2, 78.10^{-6}] \text{ ms}^{-1}$. The time T_1 spent in the workshop is drawn according to an

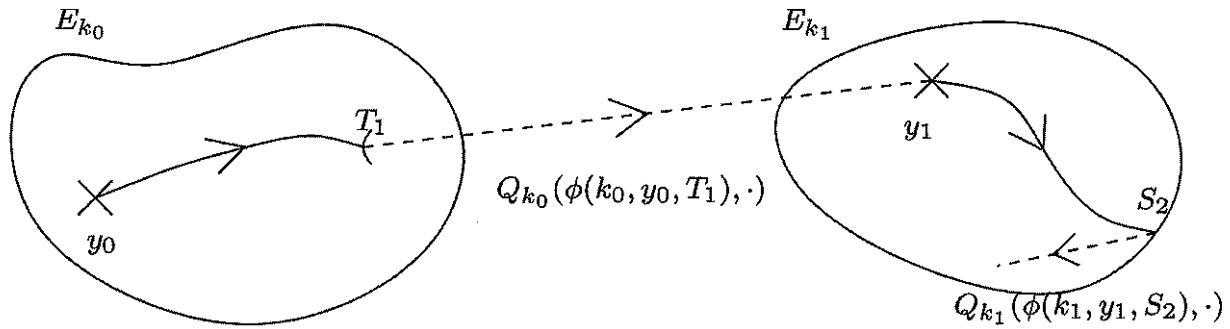


Fig. 1 An example of path for a PDMP until the second jump. The first jump is random. The second jump is deterministic because the process has reached the boundary of the domain

exponential distribution with parameter $\lambda_1 = 63, 07.10^6 \text{ s}^{-1}$. At time t between time 0 and time T_1 , the remainder of the protection is simply $\gamma_t = \max\{0, \gamma_0 - t\}$, ρ_t is constant equal to ρ_0 and the thickness loss d_t is given by

$$d_t = \begin{cases} 0 & \text{if } t \leq \gamma_0 \\ \rho_0(t - (\gamma_0 + \eta_1) + \eta_1 \exp(-\frac{t-\gamma_0}{\eta_1})) & \text{if } t > \gamma_0 \end{cases}$$

where $\eta_1 = 108.10^6 \text{ s}$.

At time T_1 , a *jump* occurs, which means there is a change of environment and a new corrosion rate is drawn for the new environment. The other two components of the process (X_t) modelling the remainder of the protection γ_t and the thickness loss d_t naturally evolve continuously. Therefore, one has $m_{T_1} = 2$, $\gamma_{T_1} = 0$ if $\gamma_0 < T_1$, $\gamma_{T_1} = \gamma_0 - T_1$ otherwise; that is to say that once the initial protection is gone, it has no effect any longer, and ρ_{T_1} is drawn according to a uniform distribution on $[2, 78.10^{-8}, 2, 78.10^{-7}] \text{ ms}^{-1}$. The process continues to evolve in the same way until the next change of environment occurring at time T_2 . Between T_1 and T_2 , we just replace ρ_0 by ρ_{T_1} , γ_0 by γ_{T_1} , η_1 by $\eta_2 = 720.10^6 \text{ s}$ and t by $t - T_1$ in equation (1). The process visits successively the three environments always in the same order: 1, 2, and 3, and then returning to environment 1. The time spent in environment i is a random variable exponentially distributed with parameters λ_i with $\lambda_1 = 63, 07.10^6 \text{ s}^{-1}$, $\lambda_2 = 47, 30.10^7 \text{ s}^{-1}$ and $\lambda_3 = 31, 54.10^6 \text{ s}^{-1}$. The thickness loss evolves continuously according to equation (1) with suitably changed parameters. The period of transition in the mode i is η_i with $\eta_1 = 108.10^6 \text{ s}$, $\eta_2 = 720.10^6 \text{ s}$ and $\eta_3 = 144.10^6 \text{ s}$. The corrosion rate ρ_i expressed in ms^{-1} is drawn at each change of environments. In environments 1 and 3, it follows a uniform distribution on $I_1 = I_3 = [2, 78.10^{-7}, 2, 78.10^{-6}]$ and in environment 2 it follows a uniform distribution on $I_2 = [2, 78.10^{-8}, 2, 78.10^{-7}]$. Therefore,

the local characteristics of the PDMP modelling this corrosion process are

- the flow Φ from $M \times (\mathbb{R}_+)^3 \times \mathbb{R}$ onto $(\mathbb{R}_+)^3$

$$\Phi(i, (d, \gamma, \rho), t) = \begin{pmatrix} [\rho(t - (\gamma + \eta_i) + \eta_i \exp(- (t - \gamma)\eta_i^{-1}))] \mathbb{1}_{\{t \geq \gamma\}} \\ (\gamma - t) \vee 0 \\ \rho \end{pmatrix}$$

- the jump intensity $\lambda(i, (d, \gamma, \rho)) = \lambda_i$
- the jump kernel

$$Q[(i, (d, \gamma, \rho); i', (d', \gamma', \rho'))] = \mathbb{1}_{\{i' = (i+1) \text{ mod } (3)\}} \mathbb{1}_{\{d = d'\}} \mathbb{1}_{\{\gamma = \gamma'\}} \mathbb{1}_{I_i}(\rho')$$

The state space $(\mathbb{R}_+)^3$ has no boundary, so t^* equals infinity. However, as only the trajectories up to the limit threshold of 2.10^{-4} m that are significant, one can set an artificial boundary, say at $d_{max} = 2.10^{-4} + 10^{-8}$. Hence, one has

$$t^*(i, (d, \gamma, \rho)) = \inf\{t > 0 : \Phi(i, (d, \gamma, \rho), t) = d_{max}\}$$

Figure 2 shows examples of simulated trajectories of the thickness loss. The slope changes correspond to changes of environment. The observed dispersion is characteristic of the random nature of the phenomenon. Note that the various physical parameters were provided by Astrium and were obtained by expert opinion.

The missile is inspected and the thickness loss of the structure under study is measured at each change of environment. Note that the structure is small enough for only one measurement point to be significant. The structure is considered unusable if the loss of thickness reaches 2.10^{-4} m . The optimal maintenance time must therefore occur before reaching this critical threshold, which could cause

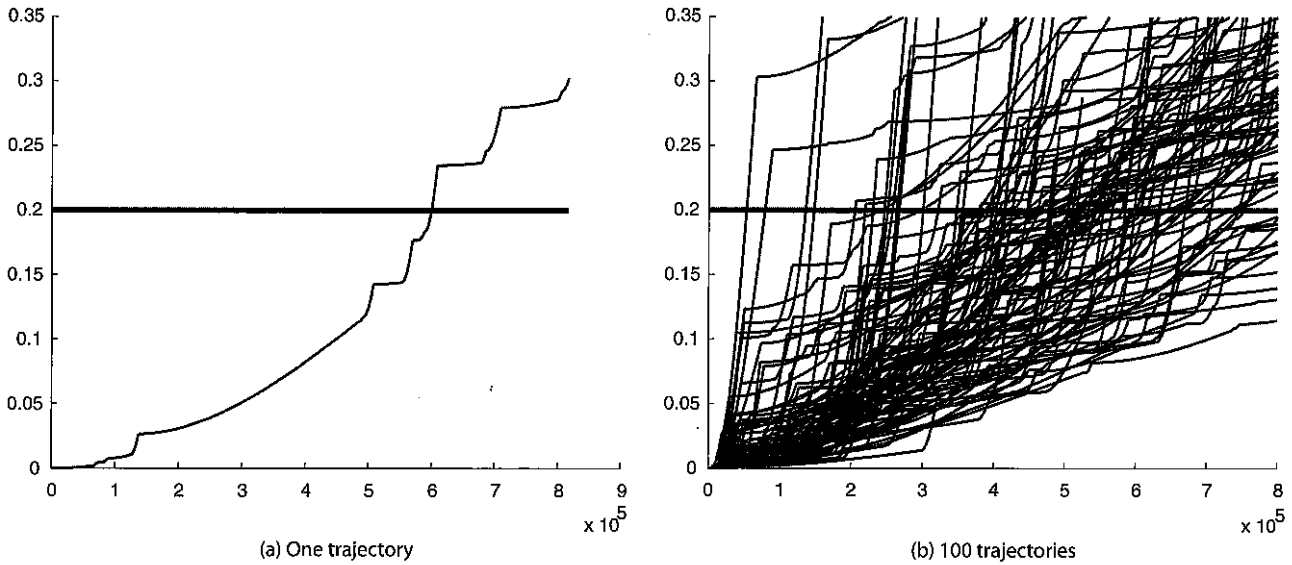


Fig. 2 Examples of trajectories of thickness loss (in mm) over time (in hours)

the collapse of the structure, but not too soon which would be unnecessarily expensive. It should also only use the available measurements of the thickness loss.

3 OPTIMAL STOPPING PROBLEM

The general mathematical problem of optimal stopping corresponding to this maintenance problem is now briefly formulated. Let $z = (k_0, y_0)$ be the starting point of the PDMP (X_t) . Let \mathcal{M}_N be the set of all stopping times τ for the natural filtration of the PDMP (X_t) satisfying $\tau \leq T_N$, that is to say that the intervention takes place before the N th jump of process. The N th jump represents the horizon of the maintenance problem, that is to say that the intervention is no later than the N th change of environment. The choice of N is discussed below. Let g be the cost function to optimize. Here, g is a reward function that has to be maximized. The optimization problem to solve is the following

$$v(z) = \sup_{\tau \in \mathcal{M}_N} E_z[g(X_\tau)]$$

The function v is called the *value function* of the problem and represents the maximum performance that can be achieved. Solving the optimal stopping problem involves first calculating the value function, and second, finding a stopping time τ that achieves this maximum. This stopping time is important from the application point of view since it corresponds to the optimum time for maintenance. In general, such an optimal stopping time does not

exist. Define then ϵ -optimal stopping times as achieving optimal value minus ϵ , that is $v(z) - \epsilon$.

Under fairly weak regularity conditions, Gugerli has shown in [9] that the value function v can be calculated iteratively as follows. Let $v_N = g$ be the reward function, and iterate an operator L backwards. The function v_0 thus obtained is equal to the value function v

$$\begin{cases} v_N = g \\ v_k = L(v_{k+1}, g), \quad 0 \leq k \leq N - 1 \end{cases}$$

The operator L is a complex operator which involves a continuous maximization, conditional expectations and indicator functions, even if the cost function g is very regular:

$$\begin{aligned} L(w, g)(z) \equiv & \sup_{u \leq t^*(z)} \{ E[w(Z_1) \mathbb{1}_{\{S_1 < u \wedge t^*(z)\}}] \\ & + g(\Phi(z, u)) \mathbb{1}_{\{S_1 \geq u \wedge t^*(z)\}} | Z_0 = z \} \\ & \vee E[w(Z_1) | Z_0 = z]. \end{aligned} \quad (2)$$

However, this operator depends only on the discrete time Markov chain (Z_n, S_n) . Gugerli also proposes an iterative construction of ϵ -optimal stopping times, which is a bit too tedious and technical to be described here, see reference [9] for details.

For the example of metallic structure, an arbitrary reward function is chosen. It depends only on the loss of thickness, since this is the critical factor to monitor. Note that the other components of our process can be taken into account without any additional difficulty. In general, the reward function should take into account several kinds of costs and constraints, namely

- the cost related to the unavailability of the structure, including any penalty;
- the intervention and maintenance or inspection cost;
- the operational constraints, an intervention cannot be scheduled at any time;
- failsafe constraints.

Their respective weight is problem-specific. In this example, the reward function is built to reflect that beyond a loss of thickness of 2.10^{-4} m, the structure is unusable, so it is too late to perform maintenance. Conversely, if the thickness loss is small, such a maintenance is unnecessarily costly. Define a piecewise affine function g for which values are given at the points in the table in Fig. 3. As for the choice of the computational horizon N , numerical simulations show that over 25 changes of environment, all trajectories exceed the critical threshold of 2.10^{-4} m. Therefore the time horizon is set to be the 25th jump ($N = 25$).

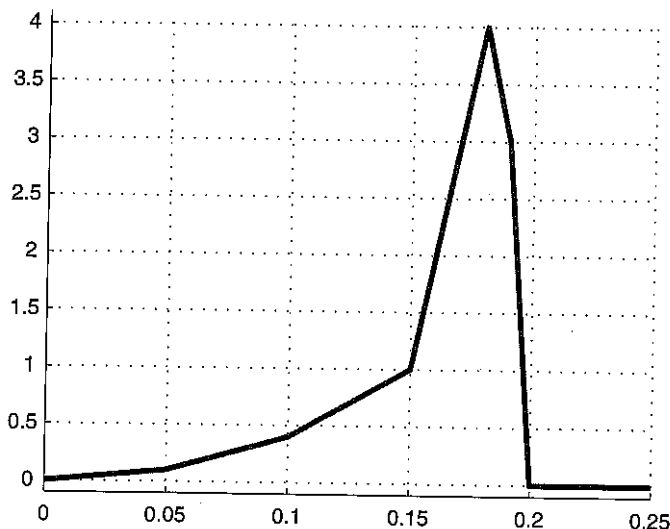
4 NUMERICAL PROCEDURE

In reference [5], the authors propose a numerical method to approximate the value function for the optimal stopping problem of a PDMP. The approach is based on quantization of the post-jump location – inter-arrival time Markov chain naturally embedded in the PDMP, and path-adapted time discretization grids. It makes possible the derivation of bounds for the convergence rate of the algorithm and provides a computable ε -optimal stopping time. The iterative algorithm proposed in reference [5] to calculate an

approximation of the value function is based on a discretization of the operator L defined in equation (2). This poses several problems, related to maximizing continuous functions, the presence of the indicator, and the presence of conditional expectations. These three problems are nevertheless overcome by using the specific properties of PDMPs, and the fact that the operator L depends only on the Markov chain (Z_n, S_n) . Our algorithm for calculating the value function is divided into three stages described below: a quantization of the Markov chain (Z_n, S_n) , a path-adapted time discretization between jumps, and finally a recursive computation of the value function v . Then, the calculation of a quasi-optimal stopping time only uses comparisons of quantities already calculated in the approximation of the value function, which makes this technique particularly attractive.

4.1 Quantization

The goal of the quantization step is to replace the continuous state space Markov chain (Z_n, S_n) by a discrete state space chain (\hat{Z}_n, \hat{S}_n) . The quantization algorithm is described in detail in, for example, references [11], [12], [13], or [14]. The principle is to obtain a finite grid adapted to the distribution of the random variable, rather than building an arbitrary regular grid. Random variables rather than the state space are discretized, the idea is to put more points in the areas of high density of the random variable. The quantization algorithm is based on Monte Carlo simulations combined with a stochastic gradient method. It provides $N + 1$ grids $\Gamma_n, 0 \leq n \leq N$ of dimension $d + 2$, one for each couple (Z_n, S_n) , with K points in each grid. The algorithm also provide



Loss of thickness (mm)	0	0.05	0.1	0.15	0.18	0.19	0.2
Value of the reward function	0	0.1	0.4	1.0	4.0	3.0	0

Fig. 3 Graphical representation and definition of the cost function as a function of the thickness loss (in mm)

weights for the grid points and probability transition between two points of two consecutive grids.

Note that p_n is the projection to the nearest neighbour (for the Euclidean norm) from \mathbb{R}^{d+2} onto Γ_n . The approximation of the Markov chain (Z_n, S_n) is constructed as

$$(\hat{Z}_n, \hat{S}_n) = p_n(Z_n, S_n)$$

Note that \hat{Z}_n and \hat{S}_n depend on both Z_n and S_n . The quantization theory ensures that the L^2 norm of the distance between (\hat{Z}_n, \hat{S}_n) and (Z_n, S_n) tends to 0 as the number of points K in the quantization grids tends to infinity, see reference [13].

It should be noted that when the dimension of Z is large, N is large, and one wants to obtain grids with a large number K of points, the quantization algorithm can be time-consuming. However, the grid calculations can be made in advance and stored. They depend only on the distribution of the process, and not on the cost function. Figure 4 gives an example of quantization grid for the standard normal distribution in two dimensions. It illustrates that the quantization algorithm puts more points in areas of high density.

4.2 Time discretization

Now the continuous maximization of the operator L is replaced by a finite maximization, that is to say that one must discretize the time intervals $[0, t^*(z)]$ for each z in the quantization grids. For this, choose a time step $\Delta < t^*(z)$ (which may depend on z) and construct the grids $G(z) = \{t_1, \dots, t_{n(z)}\}$ defined by

- $n(z)$ is the integer part minus 1 of $t^*(z)/\Delta$
- for $1 \leq i \leq n(z)$, $t_i = i\Delta$

Grids are obtained that not only do not contain $t^*(z)$, but in addition, their maximum is strictly less than $t^*(z) - \Delta$, which is a crucial property to derive error bounds for the algorithm, see reference [5]. Note also that only a finite number of grids $G(z)$ is needed, corresponding to the z in the quantization grids $(\Gamma_n)_{0 \leq n \leq N}$. Calculation of these time grids can still be made in advance. Another solution is to store only Δ and $n(z)$ which are sufficient to reconstruct the grids.

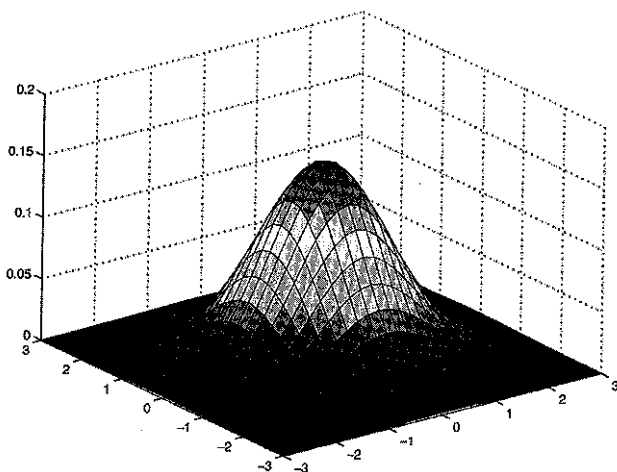
In practice, a Δ that does not depend on z is chosen. To ensure that there are no empty grids, the minimum of $t^*(z)$ on all grids of quantization is first calculated, then a Δ adapted to this value is chosen.

4.3 Approximate calculation of the value function

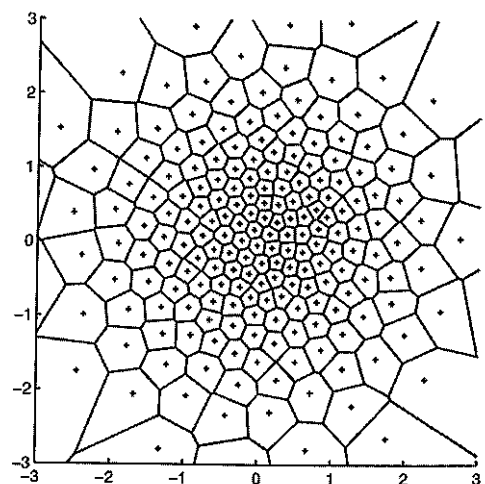
One now has all the tools to provide an approximation of the operator L . For each $1 \leq n \leq N$, and for all z in the quantization grid at time $n - 1$, set

$$\begin{aligned} \hat{L}_n(w, g)(z) = & \max_{u \in G(z)} \{ E[w(\hat{Z}_{n-1}) \mathbb{1}_{\{\hat{S}_n < u \wedge t^*(z)\}}] \\ & + g(\Phi(\hat{Z}_{n-1}, u)) \mathbb{1}_{\{\hat{S}_n \geq u \wedge t^*(z)\}} | \hat{Z}_{n-1} = z \} \\ & \vee E[w(\hat{Z}_n) | \hat{Z}_{n-1} = z] \end{aligned}$$

Note that because there are different quantized approximations at each time step, there also are different discretizations of operator L at each time step. One then constructs an approximation of the



(a) Standard normal density in 2D



(b) Quantization grid

Fig. 4 Example of quantization grid for a normal distribution

value function by backward iterations of the operators \hat{L}_n

$$\begin{cases} \hat{v}_N = g \\ \hat{v}_{n-1}(\hat{Z}_{n-1}) = \hat{L}_n(\hat{v}_n, g)(\hat{Z}_{n-1}), \quad 1 \leq n \leq N \end{cases}$$

Then take $\hat{v}_0(\hat{Z}_0) = \hat{v}_0(z)$ as an approximation of the value function v at the starting point z of the PDMP. It should be noted that the conditional expectations taken with respect to a process with discrete state space are actually finite-weighted sums.

Theorem 4.1 *Under assumptions of Lipschitz regularity of the cost function g and local characteristics (Φ, λ, Q) of the PDMP, the approximation error in the calculation of the value function is*

$$\|\hat{v}_0(z) - v_0(z)\|_2 \leq C\sqrt{EQ}$$

where C is an explicit constant which depends on the cost function and local characteristics of the PDMP, and EQ is the quantization error.

Since the quantization error tends to 0 when the number of points in the quantization grid increases, this result shows the convergence of our procedure. Here, the order of magnitude as the square root of the quantization error is due to the presence of indicator functions, which slow convergence because of their irregularity. To get around the discontinuity of these functions, it is proved that the sets where they are actually discontinuous are of very low probability. The precise statement of this theorem and its proof can be found in reference [5].

4.4 Calculation of a quasi-optimal stopping time

A method to compute an ε -optimal stopping time has also been implemented. The discretization is much more complicated and subtle than that of operator L , because one needs both to use the true Markov chain (Z_n, S_n) and its quantized version (\hat{Z}_n, \hat{S}_n) . The principle is as follows:

- At time 0, with the values $Z_0 = z$ and $S_0 = 0$, calculate a first date R_1 which depends on Z_0 , S_0 and on the value that has realized the maximum in the calculation of $\hat{L}_1(\hat{v}_1, g)$.
- Then the process is allowed to run normally until the time $R_1 \wedge T_1$, that is the minimum between this computed time R_1 and the first change of environment. If R_1 comes first, it is the date of near-optimal maintenance; if T_1 comes first, the calculation is reset.
- At time T_1 , with the values of Z_1 and S_1 , calculate the second date R_2 which depends on Z_1 and S_1

and on the the value that has realized the maximum in the calculation of $\hat{L}_2(\hat{v}_2, g)$.

- Then the process is allowed to run normally until the time $(T_1 + R_2) \wedge T_2$, that is the minimum between the computed remaining time R_2 and the next change of environment. If $T_1 + R_2$ comes first, it is the date of near-optimal maintenance; if T_2 comes first, reset the calculation, and so on until the N th jump time where maintenance will be performed if it has not occurred before.

The quality of this approximation has been proved by comparing the expectation of the cost function of the process stopped by the above strategy to the true value function. This result, its proof, and the precise construction of our stopping time procedure can be found in reference [5].

This stopping strategy is interesting for several reasons. First, this is a real stopping time for the original PDMP which is a very strong result. Second, it requires no additional computation compared to those made to approximate the value function. This procedure can be easily performed in real time, and only requires an observation of the process at the times of change of environment, which is exactly the available inspection data for our metallic structure. Moreover, even if the original problem is an optimization *on average*, this stopping rule is path-wise and is updated when new data arrive on the history of the process at each change of environment. Finally, as our stopping procedure is of the form *intervene at such date if no change of environment has occurred in the meantime*, it makes it possible in some measure to have maintenance scheduled in advance. In particular, our procedure ensures that there will be no need to perform maintenance before a given date, which is crucial for the example as a submarine in operation should not be stopped at short notice.

5 NUMERICAL RESULTS

This procedure has been implemented for the optimization of the maintenance of the metallic structure described in section 2. With the reward function chosen, it is easy to see that the true value function at our starting point is 4, which is the maximum of the reward function g , and an optimal stopping time is the first moment when the loss reaches $1.8 \cdot 10^{-4}$ m thick (value where g reaches its maximum). This is because the cost function only depends on the thickness loss, which evolves continuously increasing over time. However, our numerical procedure is valid for any sufficiently

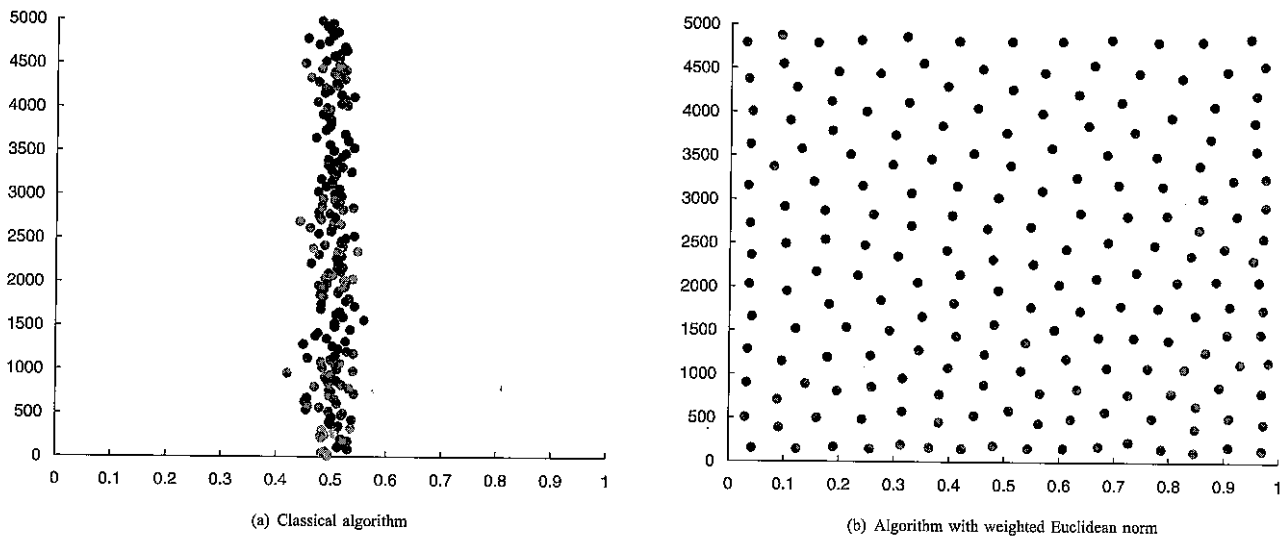


Fig. 5 Quantization grids for a uniform distribution on $[0, 1] \times [0, 5000]$

regular reward function, and the knowledge of the true value function or optimal stopping time shall not be used in the numerical procedure. Besides, recall that the thickness loss is not measured continuously.

While running the algorithm described in the previous section, an unexpected difficulty in the construction of the quantization grids was encountered. Indeed, the scales of the different variables of the problem are radically different: from about 10^{-6} for ρ to 10^5 for the average time spent in environment 2. This poses a problem in the classical quantization algorithm as searching the nearest-neighbour and gradient calculations are done in Euclidean norm, regardless of the magnitudes of the components. Figure 5 illustrates this problem by presenting two examples of quantization grids for a uniform distribution on $[0, 1] \times [0, 5000]$. The left image shows the result obtained by the conventional algorithm, the right one is obtained by weighting the Euclidean norm to renormalize each variable on the same scale. It is clear from this example that the conventional method is not satisfactory, because the grid obtained is far from uniform. This defect is corrected by a renormalization of the variables. Therefore a weighted Euclidean norm is used to quantify the Markov chain associated with our degradation process.

Figure 6 shows some projections of the quantization grids with 2000 points that were obtained. The times are chosen in order to illustrate the random and irregular nature of the grids, they are custom built to best approach the distribution of the degradation process.

Figure 7 shows two examples of computation of the quasi optimal maintenance time on two specific

simulated trajectories. The thick vertical line represents the moment provided by the algorithm to perform maintenance. The other vertical lines materialize the moments of change of environment, the horizontal dotted line the theoretical optimum. In both examples, we stop at a value very close to the optimum value. In addition, the intervention took place before the critical threshold of 2.10^{-4} m.

An approximate value function v was calculated in two ways. The first one is the direct method obtained by the algorithm described above. The second one is obtained by Monte Carlo simulation using the quasi-optimal stopping time provided by our procedure. The numerical results obtained are summarized in Table 1.

As expected, the greater the number of points in the quantization grid, the better our approximation becomes. Furthermore, the specific form of this cost function g indicates that at the threshold of 1, the intervention takes place between $1.5.10^{-4}$ and 2.10^{-4} m, and when the threshold increases, this range is narrowed. Our approximation is therefore good even for low numbers of grid points. The last column of the table also shows the validity of our stopping rule. It should be noted here that this rule does not use the optimal stopping time *stop at the first moment when the thickness loss reaches* $1.8.10^{-4}$ m. The method used is general and implementable even when the optimal stopping time is unknown or does not exist.

As already mentioned, in this particular example the optimization process has a theoretical solution since the gain function depends only on the thickness loss that evolves continuously over time. An optimal maintenance date is therefore obtained by intervening when the thickness loss reaches the

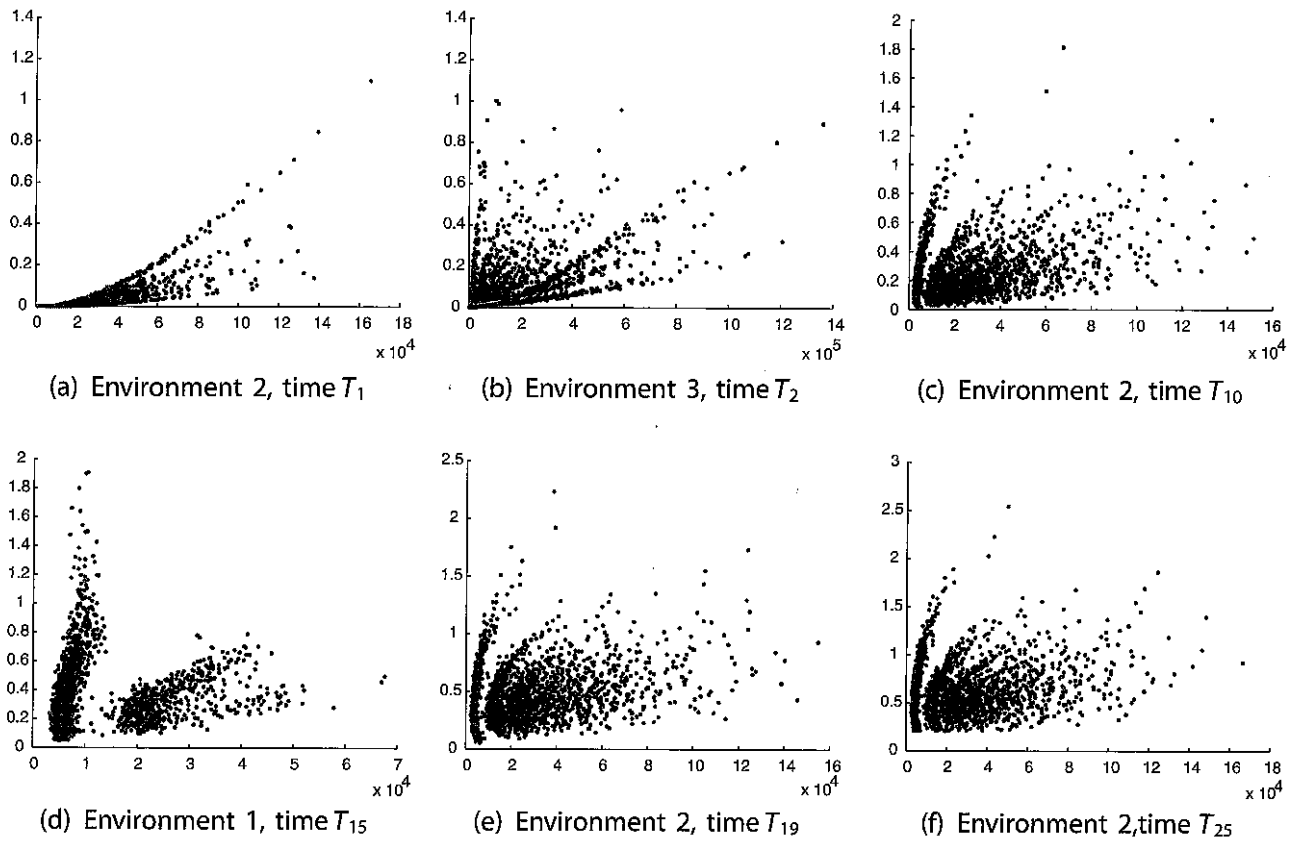


Fig. 6 Quantization grids with 2000 points for the inter-jump time (abscissa) and the thickness loss (ordinate). The scale changes for each graph

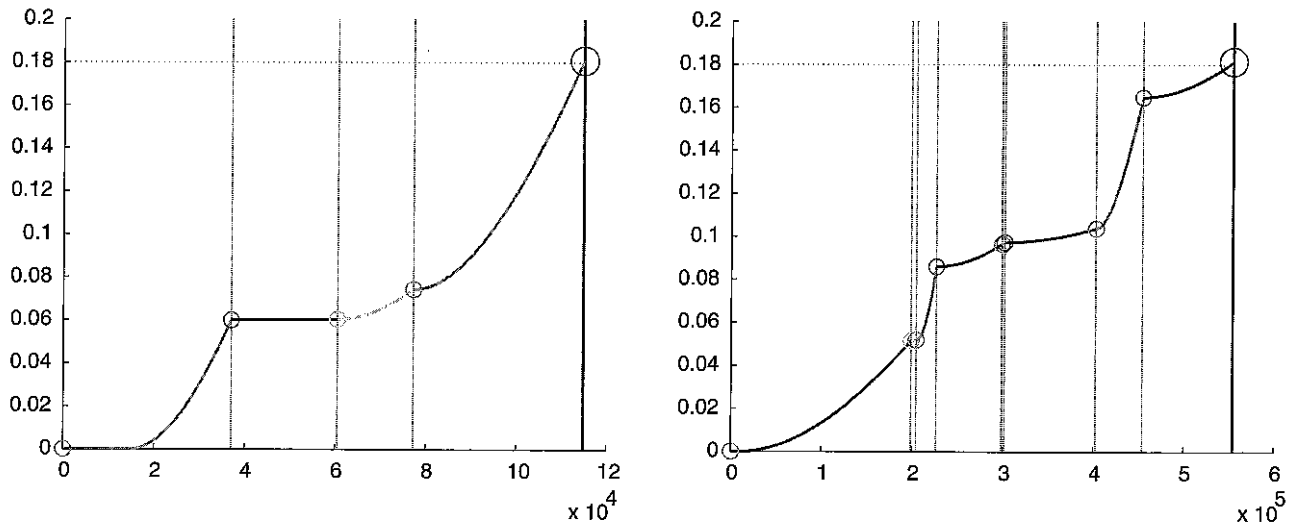


Fig. 7 Examples of stopped trajectories with the optimal maintenance time calculated by the algorithm

maximum of the gain function, that is $1,8 \cdot 10^{-4}m$. Note that the numerical procedure here does not use this theoretical solution. The maintenance date obtained by the proposed numerical procedure has been compared to the theoretical one. More

precisely, a histogram for each distribution was obtained by Monte Carlo simulations of the numerical and theoretical intervention times (Fig. 8). Both histograms are strikingly alike, showing the accuracy of our approximation. The trajectories of the

Table 1 Numerical results for the calculation of the value function

Number of points in the quantization grids	Approximation of the value function by the direct algorithm	Approximation of the value function by Monte Carlo with the quasi-optimal stopping time
10	2.48	0.94
50	2.70	1.84
100	2.94	2.10
200	3.09	2.63
500	3.39	3.15
1000	3.56	3.43
2000	3.70	3.60
5000	3.82	3.73
8000	3.86	3.75

physical corrosion process roughly speaking fall into two categories. Either one selects a high corrosion rate and so reaches the threshold early corresponding to the first peak of the distribution; or one stays a long time in a low-corrosive environment and so reaches the threshold much later, corresponding to the second peak of the distribution. This behaviour is obvious in Fig. 2 showing 100 trajectories, and in Fig. 6 showing two trends in the quantization grids, as well as in the histogram, as explained above. This point explains the bimodal distribution of the intervention time.

One can also estimate the probability that this moment is below certain thresholds, see Table 2.

These results are interesting for Astrium in the design phase of the structure to optimize margins from the specifications and to consolidate the

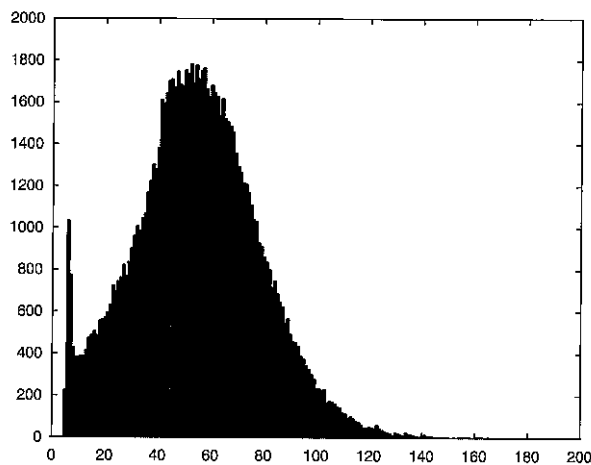
Table 2 Quantiles of the approximated intervention time

Threshold	Probability that the intervention time is below the threshold
5 years	0.0002
10 years	0.0304
15 years	0.0524
20 years	0.0793
40 years	0.2647
60 years	0.6048
80 years	0.8670
100 years	0.9691
150 years	0.9997

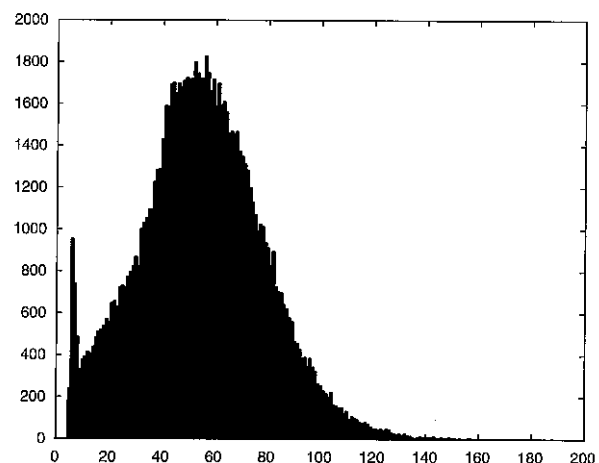
design margins available. Thus, one can justify that with a given probability no maintenance will be required before the termination date of the contract.

6 CONCLUSIONS

The numerical method described in [5] has been applied to a practical industrial example to approximate the value function of the optimal stopping problem and a quasi-optimal stopping time for a piecewise-deterministic Markov process, which is the quasi-optimal maintenance date for our structure. The quantization method proposed can sometimes be costly in computing time, but has a very interesting property: it can be calculated off-line. Moreover, it depends only on the evolutionary characteristics of the model, and not on the cost function chosen, or the actual trajectory of the specific process one wants to monitor. The calculation of



(a) Approximated.



(b) Theoretical.

Fig. 8 Distribution of the approximated (left) and theoretical (right) optimal stopping times for 100 000 Monte Carlo simulations

the optimal maintenance time is done in real time. This method is especially attractive as its application requires knowledge of the system state only at moments of change of environment and not in continuous time. The optimal maintenance time is updated at the moments when the system switches to another environment and has the form *intervene at such date if no change of mode takes place in the meantime*, which makes it possible to schedule maintenance services in advance.

This method has been implemented on an example of optimization of the maintenance of a metallic structure subject to corrosion, and very satisfactory results were obtained, very close to theoretical values, despite the relatively large size of the problem. These results are interesting for Astrium in the design phase of the structure to maximize margins from the specifications and to consolidate the available dimensional margins. Thus, tools are proposed to justify that with a given probability no maintenance will be required before the end of the contract.

The application presented here is an example of maintenance *as good as new* of the system. The next step will be to allow only partial repair of the system. The problem will then be to find simultaneously the optimal times of maintenance and optimal repair levels. Mathematically, it is an impulse control problem, the complexity of which greatly exceeds that of the optimal stopping. Here again, the problem is solved theoretically for PDMP, but there is no practical numerical method for these processes in the literature. The authors now work in this direction and hope to be able to extend the results presented above.

FUNDING

This work was partially funded by the ARPEGE program of the National Agency for Research (ANR), project FauToCoES [grant number ANR-09-SEGI-004].

© Authors 2011

REFERENCES

- 1 Aven, T. and Jensen, U. *Stochastic models in reliability*, 1999 (Springer, New York).
- 2 Meier-Hirmer, C., Riboulet, G., Sourget, F., and Roussignol, M. Maintenance optimization for a system with a gamma deterioration process and intervention delay: application to track maintenance. *Proc. IMechE, Part O: J. Risk Reliab.*, 2009, **223**(3), 189–198.
- 3 Welte, T. M. A rule-based approach for establishing states in a Markov process applied to maintenance modelling. *Proc. IMechE, Part O: J. Risk Reliab.*, 2009, **223**(1), 1–12.
- 4 Randell, D., Goldstein, M., Hardman, G., and Jonathan, P. Bayesian linear inspection planning for large-scale physical systems. *Proc. IMechE, Part O: J. Risk Reliab.*, 2010, **224**(4), 333–345.
- 5 De Saporta, B., Dufour, F., and Gonzalez, K. Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.*, 2010, **20**(5), 1607–1637.
- 6 Davis, M. *Markov models and optimization*, 1993 (Chapman and Hall, London).
- 7 Devooght, J. *Dynamic reliability*, Advances in nuclear science and technology, 1997 (Chapman and Hall, Berlin).
- 8 Dufour, F. and Dutuit, Y. Dynamic reliability : A new model. In Proceedings of *ESREL Lambda-Mu 13 Conference*, 2002, pp. 350–353.
- 9 Gugerli, U. Optimal stopping of a piecewise-deterministic Markov process. *Stochastics*, 1986, **19**, 221–236.
- 10 Costa, O. L. V. and Davis, M. H. A. Approximations for optimal stopping of piecewise-deterministic process. *Math. Control Signals Systems*, 1988, **1**(2), 123–146.
- 11 Pages, G. A space quantization method for numerical integration. *J. Comput. Appl. Math.*, 1998, **89**(1), 1–38.
- 12 Pages, G. and Pham, H. Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 2005, **11**(5), 892–932.
- 13 Pages, G., Pham, H., and Printems, J. An optimal markovian quantization algorithm for multidimensional stochastic control problem. *Stoch. Dyn.*, 2004, **4**(4), 501–545.
- 14 Pages, G., Pham, H., and Printems, J. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, 2004, pp. 253–297 (Birkhauser, Boston).

APPENDIX

List of notation

d	dimension of the state space
d_{\max}	threshold of $2 \cdot 10^{-4} + 10^{-8} \text{m}$
d_t	thickness loss at time t
∂E_m	boundary of E_m
Δ	time step
E_m	open subset of \mathbb{R}^d
η_i	transition time in environment i
$g(\cdot)$	reward function
$G(z)$	time discretization grid
γ_t	remaining anti-corrosion protection at time t
Γ_n	quantization grids

I_i	range of the corrosion rate in environment i	S_n	inter-jump times of the PDMP
K	number of discretization points	\widehat{S}_n	quantization of S_n
$L(\cdot)$	dynamic programming operator	t	time
$\widehat{L}_n(\cdot)$	approximation of L	$t^*(\cdot)$	deterministic exit time from the state space
$\lambda(\cdot)$	jumps intensity	T_n	n -th jump time of the PDMP
λ_i	mean time spent in environment i	τ	stopping time
$\Lambda(\cdot)$	integral of $\lambda \circ \Phi$ over time	$v(\cdot) = v_0(\cdot)$	value function of the optimal stopping problem
M	finite set of possible modes	$v_n(\cdot)$	iterative value function
\mathcal{M}_N	set of stopping times dominated by T_N	$\widehat{v}_n(\cdot)$	approximation of v_n
$n(z)$	number of points in $G(z)$	$w(\cdot)$	function on $M \times \mathbb{R}^d$ or on Γ_n
N	jump time horizon	$X_t = (m_t, \mathbf{x}_t)$	piecewise deterministic Markov process (PDMP)
$p_n(\cdot)$	projection operator on Γ_n	z	starting point of the PDMP
$\Phi(\cdot)$	continuous flow	Z_n	mode and location of the PDMP after the n -th jump
$Q(\cdot)$	markov kernel	\widehat{Z}_n	quantization of Z_n
R_n	sequence of tentative intervention times		
ρ_t	corrosion rate at time t		

- 2.4 [dSZ13] B. DE SAPORTA AND H. ZHANG. Predictive maintenance for the heated hold-up tank. *Reliability Engineering & System Safety*, 2013. To appear.

Author's Accepted Manuscript

Predictive maintenance for the heated hold-up tank

Benoîte de Saporta, Huilong Zhang

PII: S0951-8320(13)00051-3
DOI: <http://dx.doi.org/10.1016/j.ress.2013.02.016>
Reference: RESS4792

To appear in: *Reliability Engineering and System Safety*

Received date: 19 July 2012
Revised date: 8 February 2013
Accepted date: 17 February 2013

Cite this article as: Benoîte de Saporta and Huilong Zhang, Predictive maintenance for the heated hold-up tank, *Reliability Engineering and System Safety*, <http://dx.doi.org/10.1016/j.ress.2013.02.016>

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting galley proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



www.elsevier.com/locate/ress

Predictive maintenance for the heated hold-up tank

Benoîte de Saporta^{a,b,c}, Huilong Zhang^{b,c}

^a*Univ. Bordeaux, CNRS GREThA, UMR 5113, F-33600 Pessac, France*

^b*Univ. Bordeaux, CNRS IMB, UMR 5251, F-33400 Talence, France*

^c*INRIA Bordeaux Sud-Ouest, team CQFD, F-33400 Talence, France*

Abstract

We present a numerical method to compute an optimal maintenance date for the test case of the heated hold-up tank. The system consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps and one outlet valve. A thermal power source heats up the fluid. The failure rates of the components depends on the temperature, the position of the three components monitors the liquid level in the tank and the liquid level determines the temperature. Therefore, this system can be modeled by a hybrid process where the discrete (components) and continuous (level, temperature) parts interact in a closed loop. We model the system by a piecewise deterministic Markov process, propose and implement a numerical method to compute the optimal maintenance date to repair the components before the total failure of the system.

Keywords: Maintenance Optimization, Numerical Method, Hybrid process, Piecewise Deterministic Markov Process, Dynamic reliability

2010 MSC: 93E20, 93E25, 60J25, 34K34

1. Introduction

A complex system is inherently sensitive to failures of its components. One must therefore determine maintenance policies in order to maintain an acceptable operating condition. Optimizing the maintenance is a very important problem in the analysis of complex systems. It determines when it is best that maintenance tasks should be performed on the system in order to optimize a cost function: either maximize a performance function or conversely minimize a loss function. Moreover, this optimization must take into

account the random nature of failures and random evolution and dynamics of the system.

The example considered here is the maintenance of the heated hold-up tank, a well know test case for dynamic reliability, see e.g. [1, 2, 3, 4]. The system consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps and one outlet valve. A thermal power source heats up the fluid. The failure rate of the components depends on the temperature, the position of the three components monitors the liquid level in the tank, and in turn, the liquid level determines the temperature. The main characteristic of this system is that it can be modeled by a stochastic hybrid process, where the discrete and continuous parts interact in a closed loop. As a consequence, simulating this process and computing related reliability indices has been a challenge for the dynamic reliability community. To our best knowledge, optimization of maintenance policies for the heated hold-up tank has not been addressed yet in the literature.

The only maintenance operation considered here is the complete replacement of all the failed components and the system restarts in its initial equilibrium state. Partial repairs are not allowed. Mathematically, this problem of preventive maintenance corresponds to a stochastic optimal stopping problem as explained by example in the book of Aven and Jensen [5]. It is a difficult problem because of the closed loop interactions between the state of the components and the liquid level and temperature. A classical approach consists in using condition-based maintenance (CBM) to act on the system based on its current state and before its failure. One can for example calculate the remaining useful life (RUL) of the system and the preventive replacement is carried out when the deterioration level exceeds a certain threshold or enters in a certain state [6, 7]. Our approach also takes into account the current state of the process, but our decision rule is not based on damage accumulation nor does it correspond to hitting some threshold. Instead, it involves a performance function that reflects that the longer the system is in a functioning state the better.

The dynamics of the heated hold-up tank can be modeled by a piecewise deterministic Markov process (PDMP), see [4]. Therefore, our maintenance problem boils down to an optimal stopping problem for PDMP's. PDMP's are a class of stochastic hybrid processes that has been introduced by Davis [8] in the 80's. These processes have two components: a Euclidean component that represents the physical system (e.g. temperature, pressure, ...) and a discrete component that describes its regime of operation and/or its

environment. Starting from a state x and mode m at the initial time, the process follows a deterministic trajectory given by the laws of physics until a jump time that can be either random (e.g. it corresponds to a component failure or a change of environment) or deterministic (when a magnitude reaches a certain physical threshold, for example the pressure reaches a critical value that triggers a valve). The process restarts from a new state and a new mode of operation, and so on. This defines a Markov process. Such processes can naturally take into account the dynamic and uncertain aspects of the evolution of the system. A subclass of these processes has been introduced by Devooght [1] for an application in the nuclear field. The general model has been introduced in dynamic reliability by Dutuit and Dufour [9].

The objective and originality of this paper is twofold. First, we propose an optimization procedure for a well-known benchmark in the dynamic reliability literature. The tank model was first introduced by [12] where only one continuous variable (liquid level) is taken into account, and then in [13] and [2] where the second variable (temperature) is introduced. They have tested various Monte Carlo approaches to simulate the process to compute reliability and safety indices. In [14], the authors have used the same system to present continuous cell-to-cell mapping Markovian approach (CCCMT) still to simulate the process. The simulation of the holdup tank example has been and is still widely studied in the literature (not exhaustive) [15, 16, 17, 18, 19, 11]. Here we go one step further and not only propose to simulate the tank process but also we optimize it.

Second, even though PDMP's have been recognized as a powerful modeling tool for dynamic reliability problems [1, 9], there are very few numerical tools adapted to these processes. Our aim is to further demonstrate the high practical power of the theoretical methodology described in [10], by applying it to the tank benchmark. In [10], the authors have proposed a numerical algorithm to optimize PDMP's and have studied its theoretical properties. This optimization procedure was first applied to an example of maintenance of a metallic structure subject to corrosion, without closed loop interactions or deterministic jumps. In addition, the system has only one continuous variable and the cost function is simple and does not depend on time, see [11]. In this paper, we adapt the numerical procedure proposed in [10] to the more challenging heated hold-up tank problem with two continuous variables, deterministic jumps when these variables hit some given boundaries and closed loop interactions between continuous and discrete variables. Furthermore, we consider a cost function that depends on both continuous variables as

well as on the running time.

The remainder of this paper is organized as follows. In section 2, the dynamics of the heated hold-up tank is presented with more details as well as the framework of PDMP's. In section 3 the formulation of the optimal stopping problem for PDMP's and its theoretical solution are briefly recalled and the four main steps of the algorithm are detailed. In section 4 the numerical results obtained on the example of the tank are presented and discussed. Finally, in section 5 a conclusion and perspectives are presented.

2. Model

We are interested in the maintenance of a heated hold-up tank. The dynamics of the tank can be modeled by a piecewise deterministic Markov process (PDMP). We first describe with more details the dynamics of the tank, then we recall the definition and some basic properties of PDMP's. The tank model is a well known benchmark in dynamic reliability. It was first introduced by [12] where only one continuous variable (liquid level) is taken into account, and then in [13] and [2] where the second variable (temperature) is introduced. We have kept the values of the parameters defined in those papers.

2.1. The heated hold-up tank

The system is represented on Figure 1. It consists of a tank containing a fluid whose level is controlled by three components: two inlet pumps (units 1 and 2) and one outlet valve (unit 3). A thermal power source heats up the fluid. The variables of interest are the liquid level h , the liquid temperature θ and the state of the three components and the controller. Each component has four states: ON, OFF, Stuck ON or Stuck OFF. Once a unit is stuck (either on or off) it cannot change state. The possible transitions between these four states are given in Figure 2. Thus, by a random transition a working unit can only become stuck (either on or off). The initial state of the components is ON for units 1 and 3 and OFF for unit 2. The intensity of jumps λ^i for unit i depends on the temperature through the equation $\lambda^i = a(\theta)l^i$ with $a(\theta)$ given in Eq. (1), see [13, 2]

$$a(\theta) = \frac{b_1 \exp(b_c(\theta - 20)) + b_2 \exp(-b_d(\theta - 20))}{b_1 + b_2}. \quad (1)$$

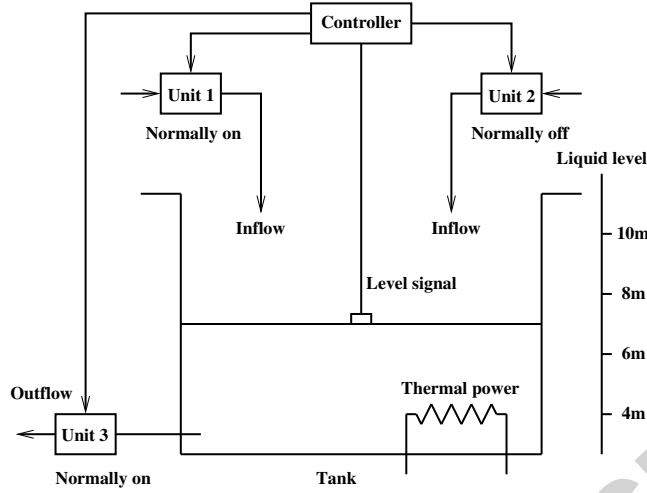
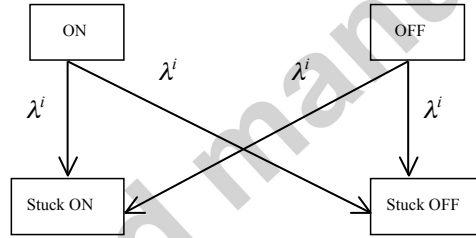
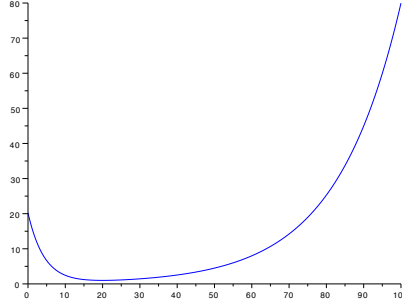


Figure 1: The heated hold-up tank

Figure 2: Transitions for unit i

Function $a(\theta)$ is represented on Figure 3 and the various parameters come from the literature, see [13, 2], and are given in Table 1. The special form of the failure rate λ^i as a product of a constant depending on i and a function of the temperature allows for all three units to have failure rates with the same dependence on the temperature, but different scaling parameters. Indeed, at the reference temperature of 20°C , the mean time to failure of unit 1 is $438h$, for unit 2, it is $350h$ and for unit 3 it is $640h$.

In addition, the shape for function $a(\theta)$ was chosen in the original benchmark so that there is a very high failure rate when the temperature is high. More specifically, the parameters are chosen such that $a(\theta)$ is lowest (equal to 1) when the temperature is equal to a reference temperature of 20°C , it equals 20 when the temperature is 0°C and it is highest (equal to 80) when

Figure 3: $a(\theta)$ as a function of θ

the temperature equals the critical temperature of 100°C . The exponential functions are chosen in order to enable this very high dependence with the temperature. Roughly speaking, the units fail 80 times more often when the temperature is 100°C than when it is 20°C .

In addition, control laws are used to modify the state of the components to keep the liquid within two acceptable limits: 6 meters and 8 meters. If the liquid level drops under 6 m, the components 1, 2, 3 are put respectively in state ON, ON and OFF (provided they are not stuck). If the liquid level rises above 8 m, the components are put respectively in the state OFF, OFF and ON (provided they are not stuck). Unlike the classical model presented in [1, 2, 3, 4], we also allow the control unit to fail. At each solicitation, the control may succeed with probability $p = 0.8$ independently from previous successes. Once it has failed, it will never succeed again. Therefore the control unit has two possible states: working 1 or failed 0.

The evolution of the liquid level h depends on the position of the three components through the differential equation (2)

$$\frac{\partial h}{\partial t} = (\nu_1 + \nu_2 - \nu_3)G, \quad (2)$$

where $\nu_i = 1$ if component i is ON or Stuck ON, $\nu_i = 0$ otherwise, and G is the common flow of the three components and is given in Table 1. The initial level is $h_0 = 7$ m. Eq. (2) simply means that each pump on contributes to rise the liquid level, whereas if the outlet valve is on, it contributes to the decrease of the liquid level. The temperature θ depends on the liquid level through the differential equation (3)

$$\frac{\partial \theta}{\partial t} = ((\nu_1 + \nu_2)G(\theta_{in} - \theta) + K)h^{-1}, \quad (3)$$

Parameter	Value
b_1	3.0295
b_2	0.7578
b_c	0.05756
b_d	0.2301
θ_{in}	15°C
l^1	$2.2831 \cdot 10^{-3} h^{-1}$
l^2	$2.8571 \cdot 10^{-3} h^{-1}$
l^3	$1.5625 \cdot 10^{-3} h^{-1}$
G	$1.5 mh^{-1}$
K	$23.88915 m^\circ Ch^{-1}$

Table 1: Parameters for the tank dynamics

where θ_{in} is the temperature of the incoming fluid, and K is a constant of the tank, the values of these parameters are given in Table 1. As the tank is heated and the incoming liquid has a constant temperature θ_{in} , Eq. (3) reflects that the temperature converges to an equilibrium state as long as there is some incoming fluid. The temperature can increase to the threshold $100^\circ C$ if there is no incoming fluid. The initial temperature is $\theta_0 = 30.9261^\circ C$, so that the system is initially at an equilibrium state, and nothing happens until the failure of one of the components. The system stops as soon as one of the top events is reached: dry out ($h < 4 m$), overflow ($h > 10 m$) or hot temperature ($\theta > 100^\circ C$).

2.2. Piecewise deterministic Markov processes

As in [4], we model the tank by a Piecewise-deterministic Markov processes (PDMP). PDMP's are a general class of hybrid processes. They are defined as follows. Let M be the finite set of the possible modes of the system. In our tank example, the modes correspond to the possible positions of the inlet pumps, outlet valve and control unit. The components can be ON, OFF, stuck ON or Stuck OFF, the control unit can be in position 0 or 1. Therefore, there are 128 possible modes for our system (but only 74 can actually be reached from the equilibrium starting point).

For all modes m in M , let E_m be an open subset in \mathbb{R}^d . In our case $d = 3$ as we need to take into account the running time as well as the liquid level and temperature, and E_m is a subset of $(4, 10) \times [15, 100) \times [0, +\infty)$ (depending

on m). A PDMP is defined from three local characteristics (Φ, λ, Q) where

- the flow $\Phi : M \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is continuous and for all $s, t \geq 0$, one has $\Phi(\cdot, \cdot, t + s) = \Phi(\Phi(\cdot, \cdot, s), t)$. It describes the deterministic trajectory of the process between jumps. In the tank example, it is given by the solution of Eq. (2) and (3). For all (m, \mathbf{x}) in $M \times E_m$, set

$$t^*(m, \mathbf{x}) = \inf\{t > 0 : \Phi(m, \mathbf{x}, t) \in \partial E_m\}, \quad (4)$$

the time to reach the boundary of the domain starting from \mathbf{x} in mode m . For the tank, the boundary is one of the thresholds $4 m, 6 m, 8 m, 10 m, 100^\circ C$ or 1000 hours for the running time.

- the jump intensity λ characterizes the frequency of jumps. For all (m, \mathbf{x}) in $M \times E_m$, and $t \leq t^*(m, \mathbf{x})$, set

$$\Lambda(m, \mathbf{x}, t) = \int_0^t \lambda(\Phi(m, \mathbf{x}, s)) ds. \quad (5)$$

For the tank the jump intensity given a mode m is the sum of the intensities λ_i for the remaining possible jumps of the three units.

- the Markov kernel Q represents the transition measure of the process and allows to select the new location and mode after each jump. In our example, Q acts only on the mode components and leaves the liquid level h , temperature θ and running time unchanged. It selects one of the remaining possible failures of the three components, or corresponds to an attempted control law.

The trajectory $\mathbf{X}_t = (m_t, \mathbf{x}_t)$ of the process can then be defined iteratively. It starts at an initial point $\mathbf{X}_0 = (k_0, y_0)$ with $k_0 \in M$ and $y_0 \in E_{k_0}$. For the tank, $k_0 = (ON, OFF, ON, 1)$ and $y_0 = (7, 30.9261, 0)$. The first jump time T_1 is determined by Eq. (6)

$$\mathbb{P}_{(k_0, y_0)}(T_1 > t) = \begin{cases} e^{-\Lambda(k_0, y_0, t)} & \text{if } t < t^*(k_0, y_0), \\ 0 & \text{if } t \geq t^*(k_0, y_0). \end{cases} \quad (6)$$

It corresponds to the first failure time of one of the components as in our case $t^*(k_0, y_0) = +\infty$. On the interval $[0, T_1)$, the process follows the deterministic trajectory $m_t = k_0$ and $\mathbf{x}_t = \Phi(k_0, y_0, t)$. At the random time T_1 , a jump

occurs. Note that in general a jump can be either a discontinuity in the Euclidean variable \mathbf{x}_t or a change of mode. The process restarts at a new mode and/or position $\mathbf{X}_{T_1} = (k_1, y_1)$, according to distribution $Q_{k_0}(\Phi(k_0, y_0, T_1), \cdot)$. An inter jump time $T_2 - T_1$ is then selected in a similar way, and on the interval $[T_1, T_2)$ the process follows the path $m_t = k_1$ and $\mathbf{x}_t = \Phi(k_1, y_1, t - T_1)$. Thereby, iteratively, a PDMP is constructed, see Figure 4 for an illustration.

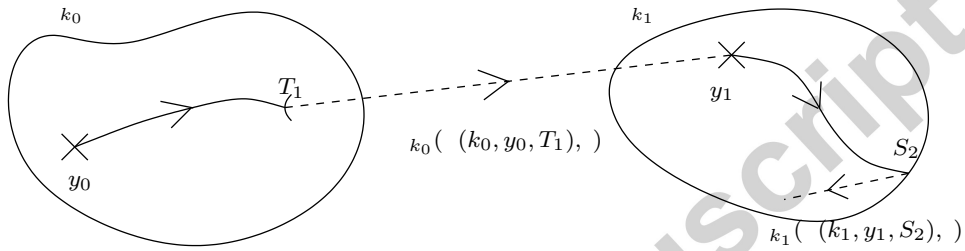


Figure 4: An example of path for a PDMP until the second jump. The first jump is random. The second jump is deterministic because the process has reached the boundary of the domain.

Let $\mathbf{Z}_0 = \mathbf{X}_0$, and for $n \geq 1$, $\mathbf{Z}_n = \mathbf{X}_{T_n}$, location and mode of the process after each jump. Let $S_0 = 0$, $S_1 = T_1$ and for $n \geq 2$, $S_n = T_n - T_{n-1}$ the inter-jump times between two consecutive jumps, then (\mathbf{Z}_n, S_n) is a Markov chain, which is the only source of randomness of the PDMP and contains all information on its random part. Indeed, if one knows the jump times and the positions after each jump, one can reconstruct the deterministic part of the trajectory between jumps. Namely, if one knows the time and nature of all the components failures, one can reconstruct the history of the liquid level and temperature though Eq. (2) and (3). It is a very important property of PDMP's that is at the basis of the numerical procedure.

3. Optimization problem

3.1. General framework

The general mathematical problem of optimal stopping corresponding to this maintenance problem can be found in [20, 10, 11]. It is now briefly recalled. Let $\mathbf{z} = (k_0, y_0)$ be the starting point of the PDMP (\mathbf{X}_t) . Let \mathcal{M}_f be the set of all stopping times τ for the natural filtration of the PDMP (\mathbf{X}_t)

satisfying $\tau \leq T_f$ that is to say that the intervention takes place before the time $T_f = 1000 h$. It has been shown in previous studies that by 1000 h , all the units are stuck and the events of interest have been observed, see e.g. [1, 2, 3, 4]. Let g be the cost function to optimize. Here, g is a reward function that has to be maximized. The optimization problem to solve is given in Eq. (7)

$$v(\mathbf{z}) = \sup_{\tau \in M_f} E_{\mathbf{z}} [g(\mathbf{X}_{\tau})]. \quad (7)$$

The function v is called the *value function* of the problem and represents the maximum performance that can be achieved. Solving the optimal stopping problem is firstly to calculate the value function, and secondly to find a stopping time τ that achieves this maximum. This stopping time is important from the application point of view since it corresponds to the optimum time for maintenance. In general, such an optimal stopping time does not exist. Define then ϵ -optimal stopping times as achieving optimal value minus ϵ , i.e. $v(\mathbf{z}) - \epsilon$.

Under fairly weak regularity conditions, Gugerli has shown in [20] that the value function v can be calculated iteratively as follows. First, choose the computational horizon N such that after N jumps, the running time t has reached T_f for almost all trajectories. Let $v_N = g$ be the reward function, and iterate an operator L backwards, see Eq. (8). The function v_0 thus obtained is equal to the value function v .

$$\begin{cases} v_N = g, \\ v_k = L(v_{k+1}, g), \quad 0 \leq k \leq N - 1. \end{cases} \quad (8)$$

Operator L defined in Eq. (9) is complex and involves a continuous maximization, conditional expectations and indicator functions, even if the reward function g is very regular:

$$\begin{aligned} L(w, g)(\mathbf{z}) & \\ & \equiv \sup_{u \leq t^*(\mathbf{z})} \left\{ E \left[w(\mathbf{Z}_1) \mathbb{1}_{\{S_1 < u \wedge t^*(\mathbf{z})\}} + g(\Phi(\mathbf{z}, u)) \mathbb{1}_{\{S_1 \geq u \wedge t^*(\mathbf{z})\}} \mid \mathbf{Z}_0 = \mathbf{z} \right] \right\} \\ & \vee E [w(\mathbf{Z}_1) \mid \mathbf{Z}_0 = \mathbf{z}]. \end{aligned} \quad (9)$$

However, this operator depends only on the discrete time Markov chain (\mathbf{Z}_n, S_n) . Gugerli also proposes an iterative construction of ϵ -optimal stopping times, which is too technical to be described here, see [20] for details.

In our example, the reward function has two components $g(h, \theta, t) = f(h, \theta)t^\alpha$. The first one f depends on the liquid level and temperature, and reflects that the reward is maximal (set to 1) when h and θ are in the normal range ($6 \text{ m} \leq h \leq 8 \text{ m}$, $\theta \leq 50^\circ\text{C}$), minimal (set to 0) when reaching the top events: dry out ($h < 4 \text{ m}$), overflow ($h > 10 \text{ m}$) or hot temperature ($\theta > 100^\circ\text{C}$) and continuous in between, see Figure 5 for an illustration. The second term t involves the time and reflects that the longer the system is functioning the higher the reward. The parameter α is set to 1.01 for smoothness.

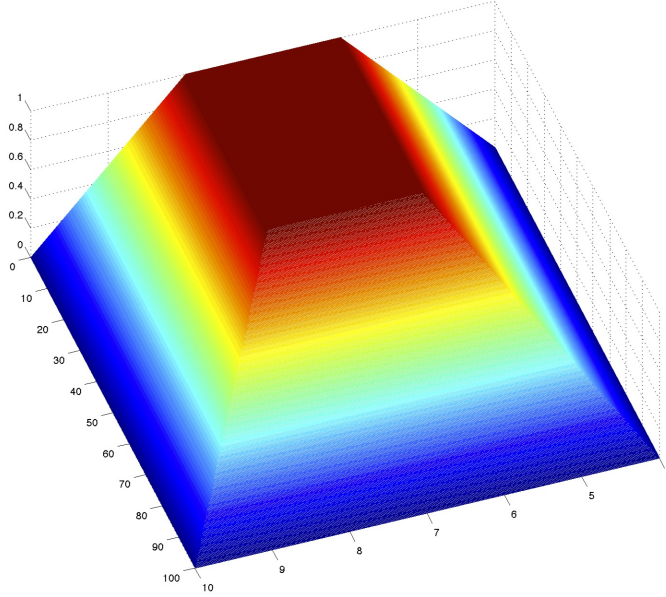


Figure 5: Reward function f as a function of h and θ

3.2. Numerical procedure

In [10, 11] the authors propose a numerical method to approximate the value function for the optimal stopping problem of a general PDMP. The approach is based on a discretization of the operator L defined above. Our algorithm for calculating the value function is divided into three stages: a discretization of the Markov chain (\mathbf{Z}_n, S_n) , a path-adapted time discretization between jumps, and finally a recursive computation of the value function v .

Then, the calculation of a quasi-optimal stopping time only uses comparisons of quantities already calculated in the approximation of the value function, which makes this technique particularly attractive. These stages are briefly recalled below.

3.2.1. Quantization

The goal of the first step is to approximate the continuous state space Markov chain (\mathbf{Z}_n, S_n) by a discrete state space sequence $(\widehat{\mathbf{Z}}_n, \widehat{S}_n)$. To this aim, we use the quantization algorithm described in details in e.g. [21, 22, 23, 24]. Roughly speaking, more points are put in the areas of high density of the random variable, see Figure 6 for an example of quantization grid for the standard normal distribution in two dimensions. The quantization

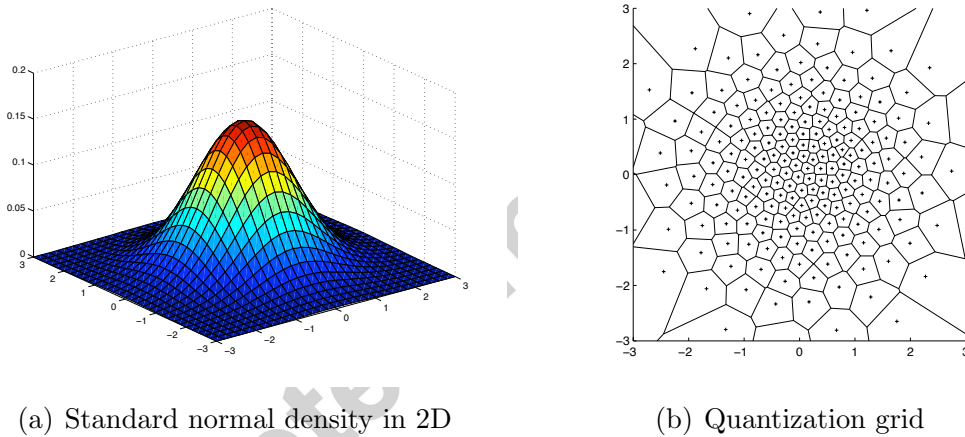


Figure 6: Example of quantization grid for a normal distribution (200 points)

algorithm is based on Monte Carlo simulations combined with a stochastic gradient method. It provides $N + 1$ grids, one for each couple (\mathbf{Z}_n, S_n) ($0 \leq n \leq N$), with a fixed number of points in each grid. The algorithm also provides weights for the grid points and probability transition between two points of two consecutive grids fully determining the distribution of the approximating sequence $(\widehat{\mathbf{Z}}_n, \widehat{S}_n)$. The quantization theory ensures that the L^2 distance between $(\widehat{\mathbf{Z}}_n, \widehat{S}_n)$ and (\mathbf{Z}_n, S_n) tends to 0 as the number of points in the quantization grids tends to infinity, see [23].

3.2.2. Time discretization

Now the continuous maximization of the operator L is replaced by a finite maximization, that is to say that one must discretize the time intervals

$[0, t^*(\mathbf{z})]$ for a finite number of \mathbf{z} , namely each \mathbf{z} in the quantization grids. For this, choose time steps $\Delta(\mathbf{z}) < t^*(\mathbf{z})$ and take a regular discretization $G(\mathbf{z})$ of $[0, t^*(\mathbf{z}) - \Delta(\mathbf{z})]$ with step $\Delta(\mathbf{z})$. The maximum in such grids is less than $t^*(\mathbf{z}) - \Delta(\mathbf{z})$, which is a crucial property to derive error bounds for the algorithm, see [10].

3.2.3. Approximate calculation of the value function

One now has all the tools to provide an approximation of the operator L given in Eq. (10). For each $1 \leq n \leq N$, and for all \mathbf{z} in the quantization grid at time $n - 1$, set

$$\begin{aligned} \widehat{L}_n(w, g)(\mathbf{z}) &\equiv \max_{u \in G(\mathbf{z})} \left\{ E \left[w(\widehat{\mathbf{Z}}_{n-1}) \mathbb{1}_{\{\widehat{S}_n < u \wedge t^*(\mathbf{z})\}} + g(\Phi(\widehat{\mathbf{Z}}_{n-1}, u)) \mathbb{1}_{\{\widehat{S}_n \geq u \wedge t^*(\mathbf{z})\}} \mid \widehat{\mathbf{Z}}_{n-1} = \mathbf{z} \right] \right\} \\ &\quad \vee E \left[w(\widehat{\mathbf{Z}}_n) \mid \widehat{\mathbf{Z}}_{n-1} = \mathbf{z} \right]. \end{aligned} \quad (10)$$

Note that because there are different quantized approximations at each time step, there also are different discretizations of operator L at each time step. It should be also noted that the conditional expectations taken with respect to a process with finite state space are actually finite weighted sums. One then constructs an approximation of the value function by backward iterations of the operators \widehat{L}_n :

$$\begin{cases} \widehat{v}_N &= g, \\ \widehat{v}_{n-1}(\widehat{\mathbf{Z}}_{n-1}) &= \widehat{L}_n(\widehat{v}_n, g)(\widehat{\mathbf{Z}}_{n-1}), \quad 1 \leq n \leq N. \end{cases} \quad (11)$$

Then take $\widehat{v}_0(\widehat{\mathbf{Z}}_0) = \widehat{v}_0(\mathbf{z})$ as an approximation of the value function v at the starting point \mathbf{z} of the PDMP. The difference between $\widehat{v}_0(\mathbf{z})$ and v goes to zero as the number of points in the quantization grids goes to infinity, see [10] for details and a convergence rate.

3.2.4. Calculation of a quasi-optimal stopping time

A method to compute an ϵ -optimal stopping time has also been implemented. The discretization is much more complicated and subtle than that of operator L , because one needs both to use the true Markov chain (\mathbf{Z}_n, S_n) and its quantized version $(\widehat{\mathbf{Z}}_n, \widehat{S}_n)$. The principle is as follows:

- At time 0, with the values $\mathbf{Z}_0 = \mathbf{z}$ and $S_0 = 0$, calculate a first date R_1 which depends on \mathbf{Z}_0 , S_0 and on the value that has realized the maximum in the calculation of $\widehat{L}_1(\widehat{v}_1, g)$.

- Then the process is allowed to run normally until the time $\min\{R_1, T_1\}$. If R_1 comes first, it is the date of maintenance, if T_1 (the date of the first failure) comes first, the calculation is reset.
- At time T_1 , with the values of \mathbf{Z}_1 and S_1 , calculate the second date R_2 which depends on \mathbf{Z}_1 and S_1 and on the the value that has realized the maximum in the calculation of $\widehat{L}_2(\widehat{v}_2, g)$.
- Then the process is allowed to run normally until the time $\min\{(T_1 + R_2), T_2\}$. If $T_1 + R_2$ comes first, it is the date of maintenance, if T_2 comes first, reset the calculation, and so on until the N th jump time or a total running time of $1000 h$, whichever comes first, where maintenance will be performed if it has not occurred before.

The quality of this approximation has been proved by comparing the expectation of the cost function of the process stopped by the above strategy to the true value function. This result, its proof and the precise construction of our stopping time procedure can be found in [10].

This stopping strategy is interesting for several reasons. First, this is a real stopping time for the original PDMP which is a very strong theoretical result. Second, it requires no additional computation compared to those made to approximate the value function. This procedure can be easily performed in real time. Moreover, even if the original problem is an optimization *on average*, this stopping rule is path-wise and is updated when new data arrive on the history of the process at each new component failure.

4. Numerical results

The numerical procedure described above is valid for a wide class of PDM-P's, see [10] for details. It has nice convergence properties. However, to implement it in practice, one must carefully choose the various parameters

- the computation horizon N ,
- the possibly state dependent time discretization steps $\Delta(\mathbf{z})$,
- the number of points in the quantization grids,
- the parameters required to build quantization grids.

These parameters may not be easy to choose for a given application. One important contribution of this paper is to show that these parameters can be suitably chosen to optimize the tank.

The numerical procedure described above has been implemented on the example of the heated holdup tank. We used the exact C++ simulator of trajectories developed for [4], suitably modified to take into account the possible failures of the command and interfaced with a matlab code for the optimization procedure. The jump horizon N was empirically set to 26 jumps, thus allowing all the trajectories to reach one of the top events $h < 4\text{ m}$, $h > 10\text{ m}$, $\theta > 100^\circ\text{C}$ or $t > 1000$ hours.

4.1. Quantization grids

We encountered a new difficulty when deriving the quantization grids, due to the high cardinality of the possible modes and possibly low probability of reaching some of them. Our mathematical model for the dynamics of the tank is hybrid: there is a discrete mode variable (the positions of the components and state of the control unit) and a continuous variable (liquid level, temperature, running time). Of course, one needs not discretize the mode variable as it can already take only finitely many values. Our procedure requires one discretization grid at each jump time of the process. However, at a given jump time, several modes can appear. For instance, at time 0, the starting mode is $(ON, OFF, ON, 1)$. After the first jump time, one of the components has failed, so there are now 6 possible modes: $(Stuck\ ON, OFF, ON, 1)$, $(stuck\ OFF, OFF, ON, 1)$, $(ON, stuck\ ON, ON, 1)$, $(ON, stuck\ OFF, ON, 1)$, $(ON, OFF, stuck\ ON, 1)$ or $(ON, OFF, stuck\ OFF, 1)$. Table 2 gives the theoretical number of possible modes at each time step as well as the observed one for $3 \cdot 10^9$ simulated trajectories. After 5 jumps, the theoretical number of modes is constant equal to 18, but all the 18 modes can actually be observed only as long as the controller unit does not fail. As the probability for the controller to remain in its operational state decreases with the number of trials of the control laws, the 18 modes become increasingly rare and by the 25-th jump time are not observed anymore, which means that the system has reached one of the top events and stopped.

The comparatively rare events are problematic for the implementation of the quantization algorithm. Indeed, one first chooses the number k of discretization points then usually initializes the algorithm by throwing k trajectories of the process at random. Thus, some rare modes may not be reached

Time	Theory	Simulations	Time	Theory	Simulation
n=0	1	1	n=20	18	17
n=1	6	6	n=21	18	16
n=2	18	18	n=22	18	14
n=3	30	30	n=23	18	7
n=4	25	25	n=24	18	1
n=5			n=25	18	0
to n=19	18	18	n=26	18	0

Table 2: Theoretical and observed number of modes at each time step for $3 \cdot 10^9$ Monte Carlo simulations.

by the initial simulations, and the algorithm will not perform well when new trajectories are thrown reaching these modes. Indeed, the algorithm is based on a nearest neighbor search, within the nodes having the same mode as the original point. When no such mode is present, the algorithm provides highly unsatisfactory results. Therefore, we had to find a way to ensure that the initializing grids have at least one point in each observed mode for each time step. To do so, at each time step we allocated the k points to the possible modes proportionally to their frequency (computed with $3 \cdot 10^9$ Monte Carlo simulations) and forcing 1 point for the modes with frequency less than $1/k$.

4.2. Optimal performance and maintenance date

We ran our optimization procedure for several number of discretization points in the quantization grids. The results we obtained are given in Table 3. The optimal performance is our approximation of the value function v at the starting equilibrium point whereas the last column gives the mean performance achieved by our stopping rule (obtained by 10^5 Monte Carlo simulations). One can observe the convergence as the number of points in the quantization grids increases, and one can also see that our stopping rule is indeed close enough to optimality.

We can also obtain the distributions of the maintenance time by Monte Carlo simulations (10^5 simulations). It is given in Figure 7. The distribution is roughly bimodal, with a very high mode (14.16% of trajectories) at time 1000, which means that the tank remained in an acceptable state until the end of the experiment. For better readability of the histogram, only the value of $\tau^* < 1000$ are plotted on the right-hand side figure. This distribution

Number of points	Optimal performance	Stopping rule
200	334.34	305.55
300	333.04	319.45
400	332.95	322.20
800	330.43	323.63
1000	330.87	324.04

Table 3: Optimal performance

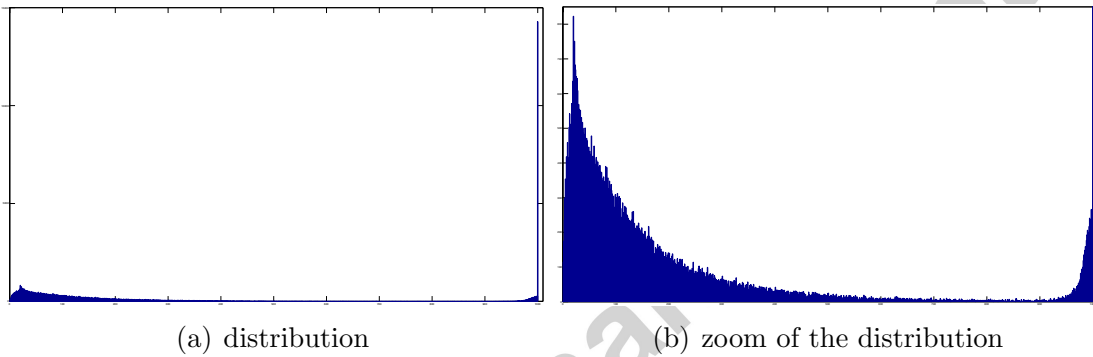


Figure 7: Histogram of the distribution of the computed maintenance time (with and without the mode at 1000)

illustrates that an average stopping rule would be far from optimal for the tank. The distribution of the liquid level and temperature at the maintenance time are given on Figure 8. The distribution of the liquid level is almost discrete with three main values at levels 6 m (15.16%), 7 m (12.68%) and 8 m (45.45%). This is natural as the liquid level is often constant, and moves very fast between these 3 states. The distribution of the temperature is also almost discrete as it is a function of the liquid level for most modes. It has a maximum at the equilibrium temperature (15,60%). Note that the top events are never reached for the liquid level, and only 0.02% of trajectories ended at $\theta = 100^\circ C$. This is a strong point in favor of our procedure.

4.3. Validation

There is no analytical solution to our optimization problem, therefore it is impossible to compare our results with the true value function. However, it must be stressed out that there is a theoretical proof in [10] that this proce-

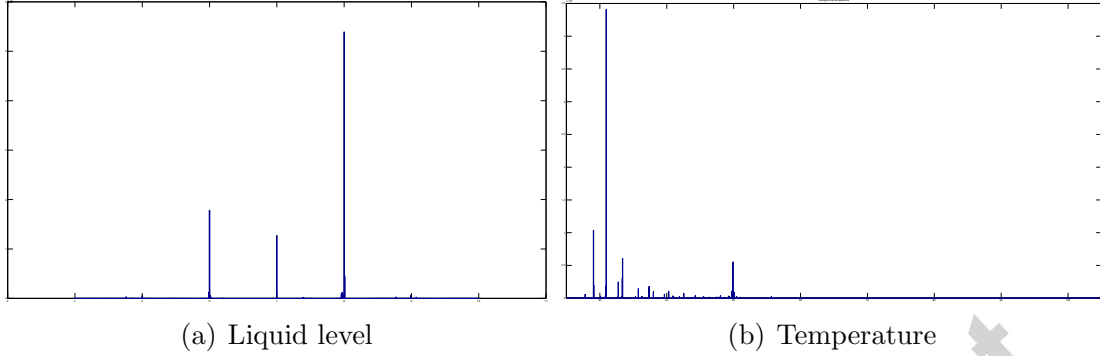


Figure 8: Histogram of the liquid level (left) and temperature (right) at the maintenance time

ture converges to the true value as the number of points in the quantization grids goes to infinity.

We can also conduct two simple kinds of experiments to validate our results. First, we can simulate new trajectories and check one by one if the intervention took place at a reasonable time. Figure 9 shows such an interesting example.

- The system starts in the equilibrium state, so that the reward function grows roughly linearly with time.
- At time 12.94 h , a jump occurs. The first unit is stuck OFF. The liquid level drops fast and the temperature rises slowly. As the liquid level and temperature are still within the acceptable bounds, the reward function is still growing roughly linearly with time.
- At time 13.61 h , the liquid level reaches the boundary 6 m and the controller switches the 3 units to stuck OFF, ON and OFF. The liquid level now rises whereas the temperature drops.
- At time 14.94 h , the liquid level reaches the boundary 8 m and the controller switches the 3 units back to stuck OFF, OFF and ON. The liquid level drops again and the temperature rises.
- At time 16.27 h , the liquid level reaches again the boundary 6 m and the controller switches the 3 units back to stuck OFF, ON and OFF. The liquid level rises and the temperature drops.

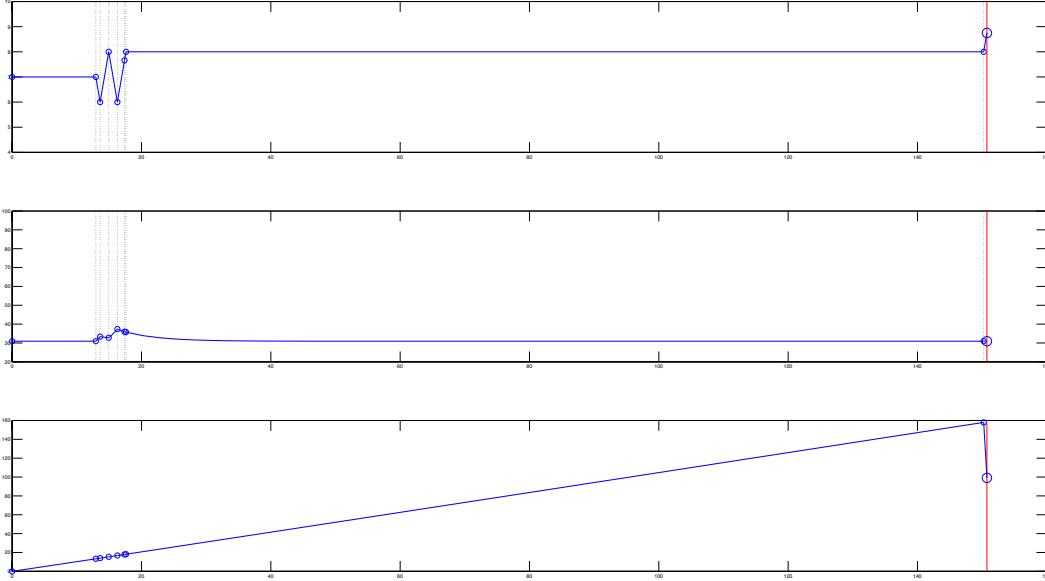


Figure 9: An example of trajectory. The upper chart shows the liquid level, the middle chart the temperature, and the bottom chart the corresponding reward, as a function of time. The small circles represent the jumps of regime, and the large circle and red line the computed maintenance time.

- At time 17.38 h , before the liquid level reaches again 8 m , the second unit fails and is now stuck ON. The liquid level still rises and the temperature drops.
- At time 17.60 h , the liquid level reaches the boundary 8 m and the controller switches the 3 units to stuck OFF, Stuck ON and ON. The liquid level now remains constant as the temperature drops down to its equilibrium state.
- At time 150.24 h , the third unit is stuck OFF. The liquid level rises again rapidly and goes above the acceptable threshold. The temperature remains constant. The algorithm decides to perform a maintenance at time 151.58. The final reward is 99.07.

Another example is given in Figure 10.

- The system starts in the equilibrium state, so that the reward function grows linearly with time.

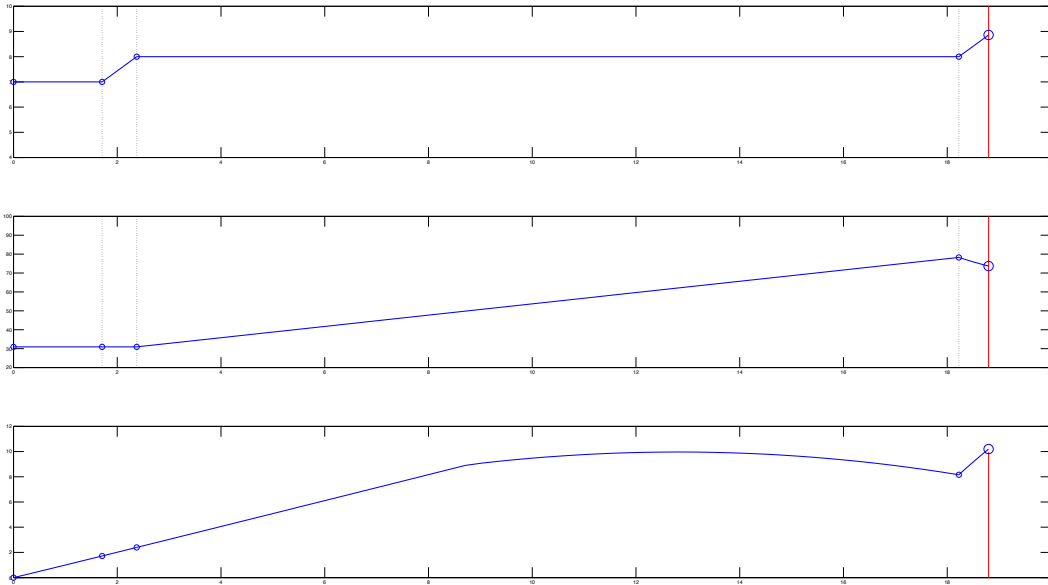


Figure 10: An example of trajectory. The upper chart shows the fluid level, the middle chart the temperature, and the bottom chart the corresponding reward, as a function of time. The small circles represent the jumps of regime, and the large circle and red line the computed maintenance time.

- At time $1.71 h$, a jump occurs. The third unit is stuck OFF. The liquid level rises and the temperature remains constant. As the liquid level and temperature are still within the acceptable bounds, the reward function is still growing linearly with time.
- At time $2.37 h$, the liquid level reaches the threshold $8m$ and triggers the command. The solicitation of the command is successful, and the first unit is turned OFF. The current state is now OFF for units 1 and 2, and stuck OFF for unit 3. The liquid level remains constant at $8 m$, but the temperature rises. At about $9 h$, the temperature crosses the threshold $50^{\circ}C$ so that the reward function is now slowly decreasing. Note that this does not trigger an immediate maintenance.
- At time $18.22 h$, another jump occurs and the second unit is now stuck ON, causing the liquid level to rise again, but the temperature to decrease. As a result, the reward function is now increasing again.
- The algorithm then selects the maintenance time to be $18.8 h$, before

	without maintenance	with maintenance
mean performance	211.80	330.87
null gain	80.33%	0.02%
$6 \leq h \leq 8$	28.25%	90.02%
$\theta \leq 50^\circ C$	80.33%	95.09%
$h = 4$	16.65%	0%
$h = 10$	54.55%	0%
$\theta = 100$	9.13%	0.02%

Table 4: Comparison of the performances with and without maintenance

the liquid level reaches the level 10 m causing a total failure of the system.

Note that at this intervention time, the first unit is not stuck, so that other jumps may happen in the future. In this special example, the intervention time occurs when the reward is maximal.

Second, one can simply compare the performances of the system when no maintenance is performed to those with our maintenance rule. The results are summarized in Table 4. These results are obtained with 10^5 Monte Carlo simulations and illustrate the power of our procedure as the mean performance is increased by 156% and the top events are almost always avoided.

5. Conclusion

The numerical method described in [10] has been applied to a well known test case of dynamic reliability to approximate the value function of the optimal stopping problem and an ϵ -optimal stopping time for a piecewise-deterministic Markov process, that is the maintenance date for the tank. The quantization method proposed can sometimes be costly in computing time, but has a very interesting property: it can be calculated off-line. Moreover it depends only on the dynamics of the model, and not on the cost function chosen, or the actual trajectory of the specific process one wants to monitor. The calculation of the optimal maintenance time is done in real time, making our procedure applicable in practice. The optimal maintenance time is updated at the changes of mode and has a conditional threshold form, which allows scheduling maintenance services in advance.

If one only changes the reward function g without changing the dynamics of the tank, one just has to run the optimization part of the algorithm, and not the quantization grids. This can be done in real time. If one wants to change the dynamics of the system, or add some components, one has to rewrite the simulation code for the system, and with this new code re-run the quantization grids, which can be quite long. However, the general methodology is valid for a wide class of piecewise deterministic Markov processes and not at all specific to the tank.

The method has been implemented on the heated hold-up tank. The main characteristic of this system is that it can be modeled by a stochastic hybrid process, where the discrete and continuous parts interact in a closed loop. The optimization problem under study has no analytic solution. However, our method is based on a rigorous mathematical construction with proof of convergence. In addition, simple comparisons between no motoring and our policy also prove its practical validity with a significant improvement of the performance of the system (the mean performance is increased by 156% and the top events are almost always avoided).

Our next project is to extend this research in two main directions. First, we could allow only partial repair of the system. The problem will then be to find simultaneously the optimal times of maintenance and optimal repair levels. Mathematically, it is an impulse control problem, which complexity exceeds widely that of the optimal stopping. Second, our method requires a perfect observation of the state process at the jump times. It would be interesting to extend our results to a noisy observation of the process, as often happens in real life.

Acknowledgements

This work was partially funded by the French National Research, ANR grant FAUTOCOES, program ARPEGE, ANR-09-SEGI-004. The authors also thank Pr. François Dufour for initial discussions on how to define an interesting optimal stopping problem for the test-case of the tank. This paper was presented at ESREL 2012 conference.

- [1] J. Devooght, Dynamic reliability, Advances in nuclear science and technology, Chapman and Hall, Berlin, 1997.

- [2] M. Marseguerra, E. Zio, The cell-to-boundary method in Monte Carlo-based dynamic PSA, *Reliability Engineering and System Safety* 48 (1995) 199–204.
- [3] M. Marseguerra, E. Zio, Monte Carlo approach to PSA for dynamic process systems, *Reliability Engineering and System Safety* 52 (1996) 227–241.
- [4] H. Zhang, F. Dufour, Y. Dutuit, K. Gonzalez, Piecewise deterministic markov processes and dynamic reliability, *Proc. of the Inst. of Mechanical Engineers, Part O: Journal of Risk and Reliability* 222 (2008) 545–551.
- [5] T. Aven, U. Jensen, Stochastic models in reliability, volume 41 of *Applications of Mathematics*, Springer, New York, 1999.
- [6] van Noortwijk J.M., A survey of the application of gamma processes in maintenance, *Reliability Engineering and System Safety* 94 (2009) 2–21.
- [7] A. Grall, C. Bérenguer, L. Dieulle, A conditional-based maintenance policy for a stochastically deteriorating system, *Reliability Engineering and System Safety* 76 (2002) 167–180.
- [8] M. H. A. Davis, Markov models and optimization, volume 49 of *Mono-graphs on Statistics and Applied Probability*, Chapman & Hall, London, 1993.
- [9] F. Dufour, Y. Dutuit, Dynamic reliability: A new model, in: *Proceedings of ESREL 2002 Lambda-Mu 13 Conference*, pp. 350–353.
- [10] B. de Saporta, F. Dufour, K. Gonzalez, Numerical method for optimal stopping of piecewise deterministic Markov processes, *Ann. Appl. Probab.* 20 (2010) 1607–1637.
- [11] B. de Saporta, F. Dufour, H. Zhang, C. Elegbede, Optimal stopping for the predictive maintenance of a structure subject to corrosion, *Proc. of the Inst. of Mechanical Engineers, Part O: Journal of Risk and Reliability* 226 (2012) 169–181.

- [12] T. Aldemir, Computer-assisted markov failure modeling of process control systems, *IEEE Transactions on Reliability* 36 (1987) 133–144.
- [13] M. Marseguerra, Approximated physical modelling in dynamic psa using artificial neural networks, *Reliability Engineering & System Safety* 45 (1994) 47–56.
- [14] B. Tombuyses, T. Aldemir, Continuous cell-to-cell mapping and dynamic psa, in: *Proceedings of ICONE 4 conference*, pp. 431–438.
- [15] N. Siu, Risk assessment for dynamic systems: an overview, *Reliability Engineering and System Safety* 43 (1994) 43–73.
- [16] G. Cojazzi, The dylam approach for dynamic reliability analysis of systems, *Reliability Engineering and System Safety* 52 (1996) 279–296.
- [17] Y. Dutuit, E. Châtelet, J. Signoret, P. Thomas, Dependability modelling and evaluation by using stochastic petri nets: application to two test cases, *Reliability Engineering and System Safety* 55 (1997) 117–124.
- [18] R. Schoenig, J. Aubry, T. Cambois, T. Hutinet, An aggregation method of markov graphs for the reliability analysis of hybrid systems, *Reliability Engineering and System Safety* 91 (2006) 137–148.
- [19] J. Li, A. Mosleh, R. Kanga, Likelihood ratio gradient estimation for dynamic reliability applications, *Reliability Engineering and System Safety* 96 (2011) 1667–1679.
- [20] U. S. Gugerli, Optimal stopping of a piecewise-deterministic Markov process, *Stochastics* 19 (1986) 221–236.
- [21] G. Pagès, A space quantization method for numerical integration, *J. Comput. Appl. Math.* 89 (1998) 1–38.
- [22] G. Pagès, H. Pham, Optimal quantization methods for nonlinear filtering with discrete-time observations, *Bernoulli* 11 (2005) 893–932.
- [23] G. Pagès, H. Pham, J. Printems, Optimal quantization methods and applications to numerical problems in finance, in: *Handbook of computational and numerical methods in finance*, Birkhäuser Boston, Boston, MA, 2004, pp. 253–297.

- [24] G. Pagès, H. Pham, J. Printems, An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems, *Stoch. Dyn.* 4 (2004) 501–545.

Accepted manuscript

- 2.5 [BdSD13] A. BRANDEJSKY, B. DE SAPORTA, AND F. DU-FOUR. Optimal stopping for partially observed piecewise-deterministic markov processes. *Stoch. Processes and their Applications*, 2013. To appear

Optimal stopping for partially observed piecewise-deterministic Markov processes*

Adrien Brandejsky Benoîte de Saporta François Dufour

Abstract

This paper deals with the optimal stopping problem under partial observation for piecewise-deterministic Markov processes. We first obtain a recursive formulation of the optimal filter process and derive the dynamic programming equation of the partially observed optimal stopping problem. Then, we propose a numerical method, based on the quantization of the discrete-time filter process and the inter-jump times, to approximate the value function and to compute an actual ϵ -optimal stopping time. We prove the convergence of the algorithms and bound the rates of convergence.

Keywords: optimal stopping, partial observation, filtering, piecewise deterministic Markov processes, quantization, numerical method
60G40, 60J25, 93E20, 93E25, 93E10, 60K10

1 Introduction

The aim of this paper is to investigate an optimal stopping problem under partial observation for piecewise-deterministic Markov processes (PDMP) both from the theoretical and numerical points of view. PDMP's have been introduced by Davis [8] as a general class of stochastic models. They form a family of Markov processes involving deterministic motion punctuated by random jumps. The motion depends on three local characteristics, the flow Φ , the jump rate λ and the transition measure Q , which selects the post-jump location. Starting from the point x , the motion of the process $(X_t)_{t \geq 0}$ follows the flow $\Phi(x, t)$ until the first jump time T_1 , which occurs either spontaneously in a Poisson-like fashion with rate $\lambda(\Phi(x, t))$ or when the flow hits the boundary of the state space. In either case, the location of the process at T_1 is selected by the transition measure $Q(\Phi(x, T_1), \cdot)$ and the motion restarts from X_{T_1} . We define similarly the time until the next jump and the next post-jump location and so on. One important property of a PDMP, relevant for the approach developed in this paper, is that its distribution is completely characterized by the discrete time Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$ where Z_n is the n -th post-jump location and S_n is the n -th inter-jump time. A suitable choice of the state space and local

*This work was supported by ARPEGE program of the French National Agency of Research (ANR), project FAUTOCOES, number ANR-09-SEGI-004.

characteristics provides stochastic models covering a large number of applications such as operations research [8, section 33], reliability [10], neurosciences [17], internet traffic [7], finance [4]. This list of examples and references is of course not exhaustive.

In this paper, we consider an optimal stopping problem for a partially observed PDMP $(X_t)_{t \geq 0}$. Roughly speaking, the observation process $(Y_t)_{t \geq 0}$ is a point process defined through the embedded discrete time Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$. The inter-arrival times are given by $(S_n)_{n \in \mathbb{N}}$ and the marks by a noisy function of $(Z_n)_{n \in \mathbb{N}}$. For a given reward function g and a computation horizon $N \in \mathbb{N}$, we study the following optimal stopping problem

$$\sup_{\sigma \leq T_N} \mathbf{E} [g(X_\sigma)],$$

where T_N is the N -th jump time of the PDMP $(X_t)_{t \geq 0}$, σ is a stopping time with respect to the natural filtration $\mathfrak{F}^Y = (\mathfrak{F}_t^Y)_{t \geq 0}$ generated by the observations $(Y_t)_{t \geq 0}$. In some applications, it may be more appropriate to consider a fixed optimization horizon t_f rather than the random horizon T_N . This is a difficult problem with few references in the literature, see for instance [11] where the underlying process is not piecewise deterministic. Regarding PDMP's, this problem could be addressed using the same ideas as in [5]. It involves the time-augmented process (X_t, t) . Although this process is still a PDMP, its local characteristics may not have the same good properties as those of the original process leading to several new technical difficulties.

A general methodology to solve such a problem is to split it into two sub-problems. The first one consists in deriving the filter process given by the conditional expectation of X_t with respect to the observed information \mathfrak{F}_t^Y . Its main objective is to transform the initial problem into a completely observed optimal stopping problem where the new state variable is the filter process. The second step consists in solving this reformulated problem, the new difficulty being its infinite dimension. Indeed, the filter process takes values in a set of probability measures.

Our work is inspired by [18] which deals with an optimal stopping problem under partial observation for a Markov chain with finite state space. The authors study the optimal filtering and convert their original problem into a standard optimal stopping problem for a continuous state space Markov chain. Then they propose a discretization method based on a quantization technique to approximate the value function. However, their method cannot be directly applied to our problem for the following main reasons related to the specificities of PDMPs.

Firstly, PDMPs are continuous time processes. Although the dynamics can be described by the discrete-time Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$, this optimization problem remains intrinsically a *continuous-time* optimization problem. Indeed, the performance criterion is maximized over the set of stopping times defined with respect to the *continuous-time* filtration $(\mathfrak{F}_t^Y)_{t \geq 0}$. Consequently, our problem cannot be converted into a fully discrete time problem.

Secondly, the distribution of a PDMP combines both absolutely continuous and singular components. This is due to the existence of forced jumps when the process hits the boundary of the state space. As a consequence the derivation of the filter process is not straightforward. In particular, the absolute continuity hypothesis **(H)** of [18] does not hold.

Thirdly, in our context the reformulated optimization problem is not standard,

unlike in [18]. As already explained, this reformulated optimization problem combines *continuous-time* and *discrete-time* features. Consequently, this problem does not correspond to the classical optimal stopping problem of a discrete-time Markov chain. Moreover, it is different from the optimal stopping problem of a PDMP under complete observation mainly because the new state variables given by the Markov chain $(\Pi_n, S_n)_{n \geq 0}$ are not the underlying Markov chain of some PDMP. Therefore the results of the literature [9, 13, 18] cannot be used.

Finally, a natural way to proceed with the numerical approximation is then to follow the ideas developed in [9, 18] namely to replace the filter Π_n and the inter-jump time S_n by some finite state space approximations in the dynamic programming equation. However, a noticeable difference from [9] lies in the fact that the dynamic programming operators therein were Lipschitz continuous whereas our new operators are only Lipschitz continuous between some points of discontinuity. We overcome this drawback by splitting the operators into their restrictions onto their continuity sets. This way, we obtain not only an approximation of the value function of the optimal stopping problem but also an ϵ -optimal stopping time with respect to the filtration $(\mathfrak{F}_t^Y)_{t \geq 0}$ that can be computed in practice.

Our approximation procedure for random variables is based on quantization. There exists an extensive literature on this method. The interested reader may for instance consult [12, 16] and the references within. The quantization of a random variable X consists in finding a finite grid such that the projection \widehat{X} of X on this grid minimizes some L^p norm of the difference $X - \widehat{X}$. Roughly speaking, such a grid will have more points in the areas of high density of X . As explained for instance in [16, section 3], under some Lipschitz-continuity conditions, bounds for the rate of convergence of functionals of the quantized process towards the original process are available, which makes this technique especially appealing. Quantization methods have been developed recently in numerical probability or optimal stochastic control with applications in finance, see e.g. [16, 2, 3].

The paper is organized as follows. Section 2 introduces the notation, recalls the definition of a PDMP, presents our assumptions and defines the optimal stopping problem we are interested in, especially the observation process. The recursive formulation of the filter process is derived in Section 3. In Section 4, we reduce our partially observed problem for the PDMP $(X_t)_{t \geq 0}$ to a completely observed one involving the process $(\Pi_n, S_n)_{n \in \mathbb{N}}$ for which we provide the dynamic programming equation and construct a family of ϵ -optimal stopping times. Then, our numerical methods to compute the value function and an ϵ -optimal stopping time are presented in Section 5 where we also prove the convergence of our algorithms after having recalled the main features of quantization. Finally, an academic example is discussed in Section 6 while technical results are postponed to the Appendices.

2 Definition and notation

In this first section, let us define a piecewise-deterministic Markov process (PDMP) and introduce some general assumptions. For any metric space E , we denote $\mathcal{B}(E)$ its Borel σ -field, $B(E)$ the set of real-valued, bounded and measurable functions defined

on E and $BL(E)$ the subset of functions of $B(E)$ that are Lipschitz continuous. For $a, b \in \mathbb{R}$, denote $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

2.1 Definition of a Piecewise-Deterministic Markov Process

Let E be an open subset of \mathbb{R}^d . Let ∂E be its boundary and \bar{E} its closure and for any subset A of E , A^c denotes its complement. A PDMP is defined by its local characteristics (Φ, λ, Q) .

- The flow $\Phi : \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathbb{R}^d$ is continuous. For all $t \in \mathbb{R}^+$, $\Phi(\cdot, t)$ is an homeomorphism and $t \rightarrow \Phi(\cdot, t)$ is a semi-group: for all $x \in \mathbb{R}^d$, $\Phi(x, t + s) = \Phi(\Phi(x, s), t)$. For all $x \in E$, define the deterministic exit time from E : $t^*(x) = \inf\{t > 0 \text{ such that } \Phi(x, t) \in \partial E\}$. We use here and throughout the convention $\inf \emptyset = +\infty$.
- The jump rate $\lambda : \bar{E} \rightarrow \mathbb{R}^+$ is measurable and satisfies:

$$\forall x \in E, \exists \epsilon > 0 \text{ such that } \int_0^\epsilon \lambda(\Phi(x, t)) dt < +\infty.$$

- Finally, Q is a Markov kernel on $(\bar{E}, \mathcal{B}(\bar{E}))$ which satisfies:

$$\forall x \in \bar{E}, Q(x, E \setminus \{x\}) = 1.$$

From these characteristics, it can be shown [8] that there exists a filtered probability space $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in \mathbb{R}^+}, (\mathbf{P}_x)_{x \in E})$ on which a process $(X_t)_{t \in \mathbb{R}^+}$ is defined. Its motion, starting from a point $x \in E$, may be constructed as follows. Let T_1 be a nonnegative random variable with survival function:

$$\mathbf{P}_x(T_1 > t) = \begin{cases} e^{-\Lambda(x, t)} & \text{if } 0 \leq t < t^*(x), \\ 0 & \text{if } t \geq t^*(x), \end{cases}$$

where for $x \in E$ and $t \in [0, t^*(x)]$, $\Lambda(x, t) = \int_0^t \lambda(\Phi(x, s)) ds$. One then chooses an E -valued random variable Z_1 with distribution $Q(\Phi(x, T_1), \cdot)$. The trajectory of X_t for $t \leq T_1$ is:

$$X_t = \begin{cases} \Phi(x, t) & \text{if } t < T_1, \\ Z_1 & \text{if } t = T_1. \end{cases}$$

Starting from the point $X_{T_1} = Z_1$, one selects in a similar way $S_2 = T_2 - T_1$ the time between T_1 and the next jump time T_2 , as well as Z_2 the next post-jump location and so on. Davis showed [8] that the process so defined is a strong Markov process $(X_t)_{t \geq 0}$ with jump times $(T_n)_{n \in \mathbb{N}}$ ($T_0 = 0$). The process $(Z_n, S_n)_{n \in \mathbb{N}}$ where $Z_n = X_{T_n}$ is the n -th post-jump location and $S_n = T_n - T_{n-1}$ ($S_0 = 0$) is the n -th inter-jump time is clearly a discrete-time Markov chain.

2.2 Notation and assumptions

The following non explosion assumption about the jump-times is standard (see for example [8, section 24]).

Assumption 2.1. For all $(x, t) \in E \times \mathbb{R}^+$, $\mathbf{E}_x \left[\sum_k \mathbb{1}_{\{T_k < t\}} \right] < +\infty$.

It implies that $T_k \rightarrow +\infty$ a.s. when $k \rightarrow +\infty$. Moreover, we make the following assumption about the transition kernel Q .

Assumption 2.2. We assume that there exists a finite set $E_0 = \{x_1, \dots, x_q\} \subset E$ such that for all $x \in E$, one has $Q(x, E_0) = 1$.

In other words, for all $n \in \mathbb{N}$, Z_n may only take its values in the finite set E_0 . This assumption ensures that the filter process, defined in the next section, has finite dimension. This is required to derive a tractable numerical method in Section 5. When this assumption does not hold, one may consider a preliminary discretization of the transition kernel to introduce it.

Assumption 2.3. We assume that the function t^* is bounded on E_0 i.e. for all $m \in \{1, \dots, q\}$, we assume that $0 < t^*(x_m) < +\infty$.

Definition 2.4. For all $m \in \{1, \dots, q\}$, denote $t_m^* = t^*(x_m)$ and assume that x_1, \dots, x_q are numbered such that $t_1^* \leq t_2^* \leq \dots \leq t_q^*$. Moreover, let $t_0^* = 0$.

For any function w in $B(E)$, introduce the following notation

$$Qw(x) = \int_E w(y)Q(x, dy) = \sum_{i=1}^q w(x_i)Q(x, x_i), \quad C_w = \sup_{x \in \bar{E}} |w(x)|.$$

For any Lipschitz continuous function w in $BL(E)$, denote $[w]$ its Lipschitz constant

$$[w] = \sup_{x \neq y \in E} \frac{|w(x) - w(y)|}{|x - y|}.$$

Assumption 2.5. The jump rate λ is in $B(\bar{E})$ i.e. is bounded by C_λ .

Denote $\mathcal{M}(E_0)$ the set of finite signed measures on E_0 and $\mathcal{M}_1(E_0)$ the subset of probability measures on E_0 . We equip $\mathcal{M}(E_0)$ with the norm $|\cdot|$ given by $|\pi| = \sum_{i=1}^q |\pi^i|$ where π^i denotes $\pi(\{x_i\})$.

2.3 Partially observed optimal stopping problem

We consider from now on a PDMP $(X_t)_{t \geq 0}$ of which the initial state $X_0 = Z_0$ is a fixed point $x_0 \in E_0$. We assume that this PDMP is observed through a noise and we now turn to the description of our observation procedure. For all $n \in \mathbb{N}$, we assume that S_n is perfectly observed but that Z_n is not (except for the initial state Z_0). In some examples, it seems reasonable to consider that the jump times of the process are observed (for instance, if the jumps correspond to changes of environment) and that, when a jump occurs, the actual post-jump location is measured with a noise. The *observation* process of Z_n , denoted by Y_n is assumed to be of the following form: $Y_0 = x_0$ (deterministic) and for $n \geq 1$,

$$Y_n = \varphi(Z_n) + W_n, \tag{1}$$

where $\varphi : E_0 \rightarrow \mathbb{R}^d$ and where the *noise* $(W_n)_{n \geq 1}$ is a sequence of \mathbb{R}^d -valued, i.i.d. random variables with bounded density function f_W that are also independent from $(Z_n, S_n)_{n \in \mathbb{N}}$. In order to define real-valued stopping times adapted to the observation process, we need to consider a continuous time version of the observation process. We therefore define the piecewise-constant process $(Y_t)_{t \geq 0}$ with a slight abuse of notation¹ as

$$Y_t = \sum_{j=0}^{+\infty} \mathbb{1}_{[T_j, T_{j+1}[}(t) Y_j.$$

Let $\mathfrak{F}^Y = (\mathfrak{F}_t^Y)_{t \geq 0}$ be the filtration generated by $(Y_t)_{t \geq 0}$ (the *observed* filtration) and $\mathfrak{F} = (\mathfrak{F}_t)_{t \geq 0}$ be the filtration generated by $(X_t, Y_t)_{t \geq 0}$ (the *total* filtration). Without changing the notation, we then complete these filtrations with all the \mathbf{P} -null sets. This leads us to the following definition.

Definition 2.6. Denote Σ^Y the set of $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times that are a.s. finite and for $n \in \mathbb{N}$, define

$$\Sigma_n^Y = \left\{ \sigma \in \Sigma^Y \text{ such that } \sigma \leq T_n \text{ a.s.} \right\}.$$

For all $n \in \mathbb{N}$, we define the filter $\Pi_n \in \mathcal{M}_1(E_0)$. The quantity $\Pi_n(\{x_i\})$, denoted by Π_n^i , represents the probability of the event $\{Z_n = x_i\}$ given the information available until time T_n i.e.

$$\forall i \in \{1, \dots, q\}, \quad \Pi_n^i = \mathbf{E}[\mathbb{1}_{\{Z_n = x_i\}} | \mathfrak{F}_{T_n}^Y]. \quad (2)$$

Finally, let $N \in \mathbb{N}$ be the *horizon* and $g \in B(\bar{E})$ the *reward function*, we are interested in maximizing the following *performance criterion*

$$\mathbf{E} \left[g(X_\sigma) | \Pi_0 = \pi \right]$$

with respect to the stopping times $\sigma \in \Sigma_N^Y$. The *value function* associated to this partially observed optimal stopping problem is given by

$$v(\pi) = \sup_{\sigma \in \Sigma_N^Y} \mathbf{E} \left[g(X_\sigma) | \Pi_0 = \pi \right], \quad (3)$$

where π is a probability measure in $\mathcal{M}_1(E_0)$. The solution of our problem is then obtained by setting $\pi = \delta_{x_0}$. For some applications, it would be interesting to consider a more general form for the reward function such as an integral term also possibly depending on the observation process, see for instance [14]. However, this new setup would lead to several technical difficulties. In particular, the dynamic programming would be more complex. Thus the derivation of the error bounds for the numerical approximation would be possibly intractable.

We will also need the following assumption about the reward function g associated with the optimal stopping problem.

¹The quantity Y_n represents the value of the process $(Y_t)_{t \geq 0}$ at time $t = T_n$ and must not be confused with the value of the process at time $t = n$.

Assumption 2.7. *The function g is in $B(\bar{E})$ i.e. bounded by C_g and there exists $[g]_2 \in \mathbb{R}^+$ such that for all $i \in \{1, \dots, q\}$ and $t, u \in [0, t_i^*]$, one has:*

$$|g(\Phi(x_i, t)) - g(\Phi(x_i, u))| \leq [g]_2 |t - u|.$$

Now, the aims of this paper are first to explicit the filter process $(\Pi_n)_{n \in \mathbb{N}}$ (Section 3); second to rewrite the partially observed optimal stopping problem (3) as a totally observed one for a suitable Markov chain on $\mathcal{M}_1(E_0) \times \mathbb{R}^+$ (Section 4.1); third to derive a dynamic programming equation and construct a family of ϵ -optimal stopping times (Section 4.2); and finally to propose a numerical method to compute an approximation of the value function and an ϵ -optimal stopping time (Section 5). As a starting point, we will derive, in the next section, a recursive construction of the optimal filter that is the key point of our approach.

3 Optimal filtering

The goal of this section is to obtain a recursive formulation of the filter Π_n . As far as we know, there is no result concerning the filter process for generic PDMPs. We may however refer to [1] for a recursive formulation of the filter for point processes, that can be seen as a sub-class of PDMP's. For all $n \in \mathbb{N}$, we denote $\mathcal{G}_n = (Y_0, S_0, \dots, Y_n, S_n)$. The continuous-time observation process $(Y_t)_{t \geq 0}$ being a point process in the sense developed in [6], one has $\mathfrak{F}_{T_n}^Y = \sigma(\mathcal{G}_n)$ (see [6, page 58, Theorem T2]). Moreover, $\mathfrak{F}_{T_n} = \sigma(Z_0, \dots, Z_n) \vee \mathfrak{F}_{T_n}^Y$. Concerning the filter Π_n , first notice that, since it is an $\mathfrak{F}_{T_n}^Y$ -measurable random variable, there exists for all $n \in \mathbb{N}$ a measurable function $\pi_n : (\mathbb{R}^d \times \mathbb{R}^+)^{n+1} \rightarrow \mathcal{M}_1(E_0)$ such that $\Pi_n = \pi_n(\mathcal{G}_n)$. As in the case of the Kalman-Bucy filter, the iteration leading from Π_{n-1} to Π_n can be split into two steps : prediction and correction. For all $n \geq 1$, let μ_n^- be the conditional distribution of (Z_n, S_n) given $\mathfrak{F}_{T_{n-1}}^Y$. Thus, μ_n^- is a transition kernel defined on $(\mathbb{R}^d \times \mathbb{R}^+)^n \times \mathcal{B}(E_0 \times \mathbb{R}^+)$ for all $j \in \{1, \dots, q\}$ and $\gamma_{n-1} \in (\mathbb{R}^d \times \mathbb{R}^+)^n$ by

$$\mu_n^-(\gamma_{n-1}, \{x_j\}, ds) = \mathbf{P}(Z_n = x_j, S_n \in ds | \mathcal{G}_{n-1} = \gamma_{n-1}). \quad (4)$$

Lemma 3.1. *For all $\gamma_{n-1} \in (\mathbb{R}^d \times \mathbb{R}^+)^n$, we have the following equality of probability measures on $E_0 \times \mathbb{R}^d \times \mathbb{R}^+$, for all $j \in \{1, \dots, q\}$,*

$$\mathbf{P}(Z_n = x_j, Y_n \in dy, S_n \in ds | \mathcal{G}_{n-1} = \gamma_{n-1}) = \mu_n^-(\gamma_{n-1}, \{x_j\}, ds) f_W(y - \varphi(x_j)) dy.$$

Proof Set h in $B(E_0 \times \mathbb{R}^d \times \mathbb{R}^+)$, using Eq. (1) that defines Y_n , one has

$$\begin{aligned} & \mathbf{E} \left[h(Z_n, Y_n, S_n) | \mathcal{G}_{n-1} = \gamma_{n-1} \right] \\ &= \sum_{j=1}^q \int h(x_j, \varphi(x_j) + w, s) \mathbf{P}(Z_n = x_j, S_n \in ds, W_n \in dw | \mathcal{G}_{n-1} = \gamma_{n-1}). \end{aligned}$$

Moreover, W_n is independent from $\sigma(Z_n, S_n) \vee \mathfrak{F}_{T_{n-1}}^Y = \sigma(Z_n, S_n, \mathcal{G}_{n-1})$ and admits the density function f_W . Consequently, one easily obtains the result by using the change of variable $y = \varphi(x_j) + w$. \square

Integrating w.r.t. to the first variable in the previous lemma (i.e. summing w.r.t. x_j) yields the following result.

Corollary 3.2. *For all $\gamma_{n-1} \in (\mathbb{R}^d \times \mathbb{R}^+)^n$, we have the following equality of probability measures on $\mathbb{R}^d \times \mathbb{R}^+$,*

$$\mathbf{P}(Y_n \in dy, S_n \in ds | \mathcal{G}_{n-1} = \gamma_{n-1}) = \left[\sum_{j=1}^q \mu_n^-(\gamma_{n-1}, \{x_j\}, ds) f_W(y - \varphi(x_j)) \right] dy.$$

Lemma 3.3. *For all $n \geq 1$, $\gamma_{n-1} \in (\mathbb{R}^d \times \mathbb{R}^+)^n$ and $j \in \{1, \dots, q\}$, the distribution μ_n^- , defined by Eq. (4), satisfies*

$$\begin{aligned} & \mu_n^-(\gamma_{n-1}, \{x_j\}, ds) \\ &= \sum_{m=0}^{q-1} \mathbb{1}_{\{s \in]t_m^*, t_{m+1}^*]\}} \left(\sum_{i=m+1}^q \pi_{n-1}^i(\gamma_{n-1}) \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} Q(\Phi(x_i, s), x_j) \right) ds \\ &+ \sum_{m=1}^q \left(\pi_{n-1}^m(\gamma_{n-1}) e^{-\Lambda(x_m, t_m^*)} Q(\Phi(x_m, t_m^*), x_j) \right) \delta_{t_m^*}(ds). \end{aligned}$$

Proof Let h be a function of $B(E_0 \times \mathbb{R}^+)$. Since $\sigma(\mathcal{G}_{n-1}) = \mathfrak{F}_{T_{n-1}}^Y \subset \mathfrak{F}_{T_{n-1}}$, the law of iterated conditional expectations yields

$$\mathbf{E} \left[h(Z_n, S_n) | \mathcal{G}_{n-1} = \gamma_{n-1} \right] = \mathbf{E} \left[\mathbf{E} \left[h(Z_n, S_n) | \mathfrak{F}_{T_{n-1}} \right] | \mathcal{G}_{n-1} = \gamma_{n-1} \right].$$

Besides, $\mathfrak{F}_{T_{n-1}} = \sigma(Z_0, S_0, W_0, \dots, Z_{n-1}, S_{n-1}, W_{n-1})$ so that

$$\mathbf{E} \left[h(Z_n, S_n) | \mathfrak{F}_{T_{n-1}} \right] = \mathbf{E} \left[h(Z_n, S_n) | Z_0, S_0, \dots, Z_{n-1}, S_{n-1} \right],$$

by independence of the sequences $(W_n)_{n \in \mathbb{N}}$ and $(Z_n, S_n)_{n \in \mathbb{N}}$. Now, we apply the Markov property of $(Z_n, S_n)_{n \in \mathbb{N}}$ and a well-known special feature of the transition kernel of the underlying Markov chain of a PDMP to obtain

$$\mathbf{E} \left[h(Z_n, S_n) | \mathfrak{F}_{T_{n-1}} \right] = \mathbf{E} \left[h(Z_n, S_n) | Z_{n-1}, S_{n-1} \right] = \mathbf{E} \left[h(Z_n, S_n) | Z_{n-1} \right].$$

Moreover, the transition kernel can be explicitly expressed in terms of the local characteristics of the PDMP, and this yields the next equations

$$\begin{aligned} & \mathbf{E}[h(Z_n, S_n) | \mathcal{G}_{n-1} = \gamma_{n-1}] \\ &= \mathbf{E} \left[\sum_{i=1}^q \mathbb{1}_{\{Z_{n-1} = x_i\}} \mathbf{E}[h(Z_n, S_n) | Z_{n-1} = x_i] | \mathcal{G}_{n-1} = \gamma_{n-1} \right] \\ &= \mathbf{E} \left[\sum_{i=1}^q \mathbb{1}_{\{Z_{n-1} = x_i\}} \sum_{j=1}^q \left[\int_{\mathbb{R}^+} h(x_j, s) \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} \mathbb{1}_{\{s < t_i^*\}} Q(\Phi(x_i, s), x_j) ds \right. \right. \\ &\quad \left. \left. + h(x_j, t_i^*) e^{-\Lambda(x_i, t_i^*)} Q(\Phi(x_i, t_i^*), x_j) \right] | \mathcal{G}_{n-1} = \gamma_{n-1} \right] \\ &= \sum_{j=1}^q \left(\int_{\mathbb{R}^+} h(x_j, s) \sum_{i=1}^q \pi_{n-1}^i(\gamma_{n-1}) \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} \mathbb{1}_{\{s < t_i^*\}} Q(\Phi(x_i, s), x_j) ds \right. \\ &\quad \left. + \sum_{i=1}^q h(x_j, t_i^*) \pi_{n-1}^i(\gamma_{n-1}) e^{-\Lambda(x_i, t_i^*)} Q(\Phi(x_i, t_i^*), x_j) \right). \end{aligned}$$

This can be written equivalently as

$$\begin{aligned} & \mathbf{E} \left[h(Z_n, S_n) \middle| \mathcal{G}_{n-1} = \gamma_{n-1} \right] \\ &= \sum_{j=1}^q \left(\sum_{m=0}^{q-1} \left(\int_{t_m^*}^{t_{m+1}^*} h(x_j, s) \sum_{i=m+1}^q \pi_{n-1}^i(\gamma_{n-1}) \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} Q(\Phi(x_i, s), x_j) \right) ds \right. \\ & \quad \left. + \sum_{i=1}^q h(x_j, t_i^*) \pi_{n-1}^i(\gamma_{n-1}) e^{-\Lambda(x_i, t_i^*)} Q(\Phi(x_i, t_i^*), x_j) \right). \end{aligned}$$

Hence the result. \square

We now state the main result of this section, namely the recursive formulation of the filter sequence $(\Pi_n)_{n \in \mathbb{N}}$.

Proposition 3.4. *Let $\Psi = (\Psi^1, \dots, \Psi^q) : \mathcal{M}_1(E_0) \times \mathbb{R}^d \times \mathbb{R}^+ \rightarrow \mathcal{M}_1(E_0)$ be defined as follows: for all $j \in \{1, \dots, q\}$,*

$$\Psi^j(\pi, y, s) = \sum_{m=0}^{q-1} \mathbb{1}_{\{s \in]t_m^*, t_{m+1}^*[\}} \frac{\Psi_m^j(\pi, y, s)}{\bar{\Psi}_m(\pi, y, s)} + \sum_{m=1}^q \mathbb{1}_{\{s=t_m^*\}} \frac{\Psi_m^{*j}(y)}{\bar{\Psi}_m^*(y)},$$

where

$$\begin{aligned} \Psi_m^j(\pi, y, s) &= \sum_{i=m+1}^q \pi^i \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} Q(\Phi(x_i, s), x_j) f_W(y - \varphi(x_j)), \\ \bar{\Psi}_m(\pi, y, s) &= \sum_{k=1}^q \Psi_m^k(\pi, y, s), \\ \Psi_m^{*j}(y) &= Q(\Phi(x_m, t_m^*), x_j) f_W(y - \varphi(x_j)), \\ \bar{\Psi}_m^*(y) &= \sum_{k=1}^q \Psi_m^{*k}(y). \end{aligned}$$

Then, the filter, defined in Eq. (2), satisfies $\Pi_0^j = \mathbf{P}(Z_0 = x_j)$ and the following recursion: for all $n \geq 1$,

$$\mathbf{P}\text{-a.s.}, P i_n = \Psi(\Pi_{n-1}, Y_n, S_{-n}).$$

Proof Fix γ_{n-1} in $(\mathbb{R}^d \times \mathbb{R}^+)^n$. Bayes formula yields for all $j \in \{1, \dots, q\}$,

$$\begin{aligned} & \mathbf{P}(Z_n = x_j, Y_n \in dy, S_n \in ds \middle| \mathcal{G}_{n-1} = \gamma_{n-1}) = \\ & \quad \mathbf{P}(Z_n = x_j \middle| \mathcal{G}_n = (\gamma_{n-1}, y, s)) \times \mathbf{P}(Y_n \in dy, S_n \in ds \middle| \mathcal{G}_{n-1} = \gamma_{n-1}). \end{aligned}$$

Lemma 3.1 and Corollary 3.2 yield

$$\begin{aligned} & \mu_n^-(\gamma_{n-1}, \{x_j\}, ds) f_W(y - \varphi(x_j)) dy \\ &= \mathbf{P}(Z_n = x_j \middle| \mathcal{G}_n = (\gamma_{n-1}, y, s)) \left[\sum_{k=1}^q \mu_n^-(\gamma_{n-1}, \{x_k\}, ds) f_W(y - \varphi(x_k)) \right] dy. \end{aligned}$$

With respect to y , one recognizes the equality of two absolutely continuous measures which implies the equality a.e. of the density functions. Thus, one has for almost all $y \in \mathbb{R}^d$ w.r.t. the Lebesgue measure,

$$\begin{aligned} & \mu_n^-(\gamma_{n-1}, \{x_j\}, ds) f_W(y - \varphi(x_j)) \\ &= \mathbf{P} \left(Z_n = x_j \mid \mathcal{G}_n = (\gamma_{n-1}, y, s) \right) \left[\sum_{k=1}^q \mu_n^-(\gamma_{n-1}, \{x_k\}, ds) f_W(y - \varphi(x_k)) \right]. \end{aligned} \quad (5)$$

Eq. (5) states the equality of two measures of the variable $s \in \mathbb{R}^+$ that contain both an absolutely continuous part and some weighted Dirac measures. Denote $g_1(y, s)\nu_1(ds)$ (respectively $g_2(y, s)\nu_2(ds)$) the left-hand (resp. right-hand) side term of the previous equality. Eq. (5) means that for all function $F \in B(\mathbb{R}^+)$ and for almost all $y \in \mathbb{R}^d$ w.r.t. the Lebesgue measure, one has

$$\int F(s)g_1(y, s)\nu_1(ds) = \int F(s)g_2(y, s)\nu_2(ds), \quad (6)$$

Recall that, from Lemma 3.3, the distribution $\mu_n^-(\gamma_{n-1}, \{x_j\}, ds)$ has a density on the interval $]t_m^*; t_{m+1}^*[$ denoted by $f_m(\gamma_{n-1}, x_j, s)$ and given by

$$f_m(\gamma_{n-1}, x_j, s) = \sum_{i=m+1}^q \pi_{n-1}^i(\gamma_{n-1}) \lambda(\Phi(x_i, s)) e^{-\Lambda(x_i, s)} Q(\Phi(x_i, s), x_j).$$

First, take $F(s) = H(s)\mathbb{1}_{\{s \in]t_m^*; t_{m+1}^*[\}}$ in equation (6) with $H \in B(\mathbb{R}^+)$. One has from equation (5)

$$\begin{aligned} & \int_{t_m^*}^{t_{m+1}^*} H(s) f_m(\gamma_{n-1}, x_j, s) f_W(y - \varphi(x_j)) ds \\ &= \int_{t_m^*}^{t_{m+1}^*} H(s) \mathbf{P} \left(Z_n = x_j \mid \mathcal{G}_n = (\gamma_{n-1}, y, s) \right) \sum_{k=1}^q f_m(\gamma_{n-1}, x_k, s) f_W(y - \varphi(x_k)) ds, \end{aligned}$$

and thus on $]t_m^*; t_{m+1}^*[$, almost surely w.r.t. the Lebesgue measure, one has

$$\mathbf{P} \left(Z_n = x_j \mid \mathcal{G}_n = (\gamma_{n-1}, y, s) \right) = \frac{f_m(\gamma_{n-1}, x_j, s) f_W(y - \varphi(x_j))}{\sum_{k=1}^q f_m(\gamma_{n-1}, x_k, s) f_W(y - \varphi(x_k))}.$$

Finally, for $m \in \{1, \dots, q\}$, choosing $F(s) = \mathbb{1}_{\{s=t_m^*\}}$ in Eq. (6) yields the equality of the weights at the point t_m^* thus, using Lemma 3.3,

$$\begin{aligned} & \mathbf{P} \left(Z_n = x_j \mid \mathcal{G}_n = (\gamma_{n-1}, y, t_m^*) \right) \\ &= \frac{\pi_{n-1}^m(\gamma_{n-1}) e^{-\Lambda(x_m, t_m^*)} Q(\Phi(x_m, t_m^*), x_j) f_W(y - \varphi(x_j))}{\sum_{k=1}^q \pi_{n-1}^m(\gamma_{n-1}) e^{-\Lambda(x_m, t_m^*)} Q(\Phi(x_m, t_m^*), x_k) f_W(y - \varphi(x_k))} \\ &= \frac{Q(\Phi(x_m, t_m^*), x_j) f_W(y - \varphi(x_j))}{\sum_{k=1}^q Q(\Phi(x_m, t_m^*), x_k) f_W(y - \varphi(x_k))}. \end{aligned}$$

Thus there exists two measurable sets $N_y \subset \mathbb{R}^d$ and $N_s \subset \mathbb{R}^+ \setminus \{t_1^*, \dots, t_q^*\}$, negligible w.r.t. the Lebesgue measures on \mathbb{R}^d and \mathbb{R} respectively, such that for all $\gamma_{n-1} \in (\mathbb{R}^d \times \mathbb{R}^+)^n$, $y \in \mathbb{R}^d \setminus N_y$, $s \in \mathbb{R}^+ \setminus N_s$, one has

$$\pi_n(\gamma_{n-1}, y, s) = \Psi(\pi_{n-1}(\gamma_{n-1}), y, s). \quad (7)$$

On the one hand, one has $\mathbf{P}(Y_n \in N_y) \leq \sum_{j=1}^q \mathbf{P}(\varphi(x_j) + W_n \in N_y) = 0$ by absolute continuity of the distribution of W_n . On the other hand, $\mathbf{P}(S_n \in N_s) = 0$ because the distribution of S_n is absolutely continuous on $\mathbb{R}^+ \setminus \{t_1^*, \dots, t_q^*\}$ and one has $N_s \cap \{t_1^*, \dots, t_q^*\} = \emptyset$. We therefore conclude from Eq. (7) that \mathbf{P} -a.s., one has $\pi_n(\mathcal{G}_{n-1}, Y_n, S_n) = \Psi(\pi_{n-1}(\mathcal{G}_{n-1}), Y_n, S_n)$. The result follows since \mathbf{P} -a.s., one has $\pi_n(\mathcal{G}_{n-1}, Y_n, S_n) = \Pi_n$ and $\pi_{n-1}(\mathcal{G}_{n-1}) = \Pi_{n-1}$. \square

This proposition will play a crucial part in the sequel. On the one hand, this result will enable us to prove the Markov property of the sequence $(\Pi_n, S_n)_{n \geq 0}$ w.r.t. the observed filtration. On the other hand, the recursive formulation allows for simulation of the process $(\Pi_n)_{n \geq 0}$ which is crucial to obtain numerical approximations. Finally, notice that the specific structure of the PDMP appears in the recursive formulation of the filter which contains both an absolutely continuous part and some weighted points.

4 Dynamic programming

The main objective of this section is to derive the dynamic programming equation for the value function of the partially observed optimal stopping problem (3). The proof of this result can be roughly speaking decomposed into two steps. The first point consists in converting the partially observed optimal stopping problem into an optimal stopping problem under complete observation where the state variables are described by the *discrete-time* Markov chain $(\Pi_n, S_n)_{n \geq 0}$ (see Section 4.1). It is important to remark that under this new formulation, the optimization problem remains intrinsically a *continuous-time* optimization problem because the performance criterion is maximized over the set of stopping times with respect to the *continuous-time* filtration $(\mathfrak{F}_t^Y)_{t \geq 0}$. We show in the second step (see Section 4.2) that the value function associated to the optimal stopping problem (3) can be calculated by iterating a functional operator, labelled L (see Definition 4.3). As a by-product, we also provide a family of ϵ -optimal stopping times.

We would like to emphasize that the results obtained in this section are not straightforward to obtain due to the specific structure of this optimization problem. Indeed, as already explained, it combines *continuous-time* and *discrete-time* features. Consequently, this problem does not correspond to the classical optimal stopping problem of a discrete-time Markov chain. Moreover, it is different from the optimal stopping problem of a PDMP under complete observation mainly because the new state variables given by the Markov chain $(\Pi_n, S_n)_{n \geq 0}$ are not the underlying Markov chain of some PDMP. Therefore the results of the literature [9, 13] cannot be used.

These derivations require some technical results about the structure of the stopping times in Σ_N^Y . For the sake of clarity in exposition, they are presented in the Appendix A. We start with a technical preliminary result required in the sequel, investigating the Markov property of the filter process.

Proposition 4.1. *The sequences $(\Pi_n, Y_n, S_n)_{n \in \mathbb{N}}$, $(\Pi_n, S_n)_{n \in \mathbb{N}}$ and $(\Pi_n)_{n \in \mathbb{N}}$ are $(\mathfrak{F}_{T_n}^Y)_{n \in \mathbb{N}}$ -Markov chains.*

Proof Let $h \in B(\mathcal{M}_1(E_0) \times \mathbb{R}^d \times \mathbb{R}^+)$. The law of iterated conditional expectations yields

$$\mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}^Y] = \mathbf{E}\left[\mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}] \middle| \mathfrak{F}_{T_{n-1}}^Y\right].$$

From Proposition 3.4 and Eq. (1) which defines Y_n one obtains

$$\begin{aligned} & \mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}] \\ &= \mathbf{E}\left[h\left(\Psi(\Pi_{n-1}, \varphi(Z_n) + W_n, S_n), \varphi(Z_n) + W_n, S_n\right) \middle| \mathfrak{F}_{T_{n-1}}\right] \\ &= \sum_{j=1}^q \int h\left(\Psi(\Pi_{n-1}, \varphi(x_j) + w, s), \varphi(x_j) + w, s\right) \\ & \quad \times \mathbf{P}(Z_n = x_j, W_n \in dw, S_n \in ds | \mathfrak{F}_{T_{n-1}}). \end{aligned}$$

Yet, W_n is independent from $\sigma(Z_n, S_n) \vee \mathfrak{F}_{T_{n-1}}$ and admits the density function f_W . As in the proof of Lemma 3.1 one thus obtains

$$\begin{aligned} & \mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}] \\ &= \sum_{j=1}^q \int h\left(\Psi(\Pi_{n-1}, y, s), y, s\right) \mathbf{P}(Z_n = x_j, S_n \in ds | \mathfrak{F}_{T_{n-1}}) f_W(y - \varphi(x_j)) dy. \end{aligned}$$

Besides, we have $\mathbf{P}(Z_n = x_j, S_n \in ds | \mathfrak{F}_{T_{n-1}}) = \mathbf{P}(Z_n = x_j, S_n \in ds | Z_{n-1})$ as in the proof of Lemma 3.3, so that one has

$$\begin{aligned} & \mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}] \\ &= \sum_{i=1}^q \mathbb{1}_{\{Z_{n-1} = x_i\}} \sum_{j=1}^q \int \left(\int_0^{t_i^*} h\left(\Psi(\Pi_{n-1}, y, s), y, s\right) \lambda\left(\Phi(x_i, s)\right) e^{-\Lambda(x_i, s)} Q\left(\Phi(x_i, s), x_j\right) ds \right. \\ & \quad \left. + h\left(\Psi(\Pi_{n-1}, y, t_i^*), y, t_i^*\right) e^{-\Lambda(x_i, t_i^*)} Q\left(\Phi(x_i, t_i^*), x_j\right) \right) f_W(y - \varphi(x_j)) dy. \end{aligned}$$

Take now the conditional expectation w.r.t. $\mathfrak{F}_{T_{n-1}}^Y$, to obtain

$$\begin{aligned} & \mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}] \\ &= \sum_{i=1}^q \Pi_{n-1}^i \sum_{j=1}^q \int \left(\int_0^{t_i^*} h\left(\Psi(\Pi_{n-1}, y, s), y, s\right) \lambda\left(\Phi(x_i, s)\right) e^{-\Lambda(x_i, s)} Q\left(\Phi(x_i, s), x_j\right) ds \right. \\ & \quad \left. + h\left(\Psi(\Pi_{n-1}, y, t_i^*), y, t_i^*\right) e^{-\Lambda(x_i, t_i^*)} Q\left(\Phi(x_i, t_i^*), x_j\right) \right) f_W(y - \varphi(x_j)) dy. \end{aligned}$$

Hence $\mathbf{E}[h(\Pi_n, Y_n, S_n) | \mathfrak{F}_{T_{n-1}}^Y]$ is merely a function of Π_{n-1} yielding the result for the three processes. \square

4.1 Optimal stopping problem under complete observation

In this section, we show how our optimal stopping problem under partial observation for the process $(X_t)_{t \geq 0}$ can be converted into an optimal stopping problem under complete observation involving the Markov chain $(\Pi_n, S_n)_{0 \leq n \leq N}$. More precisely, for

a fixed stopping time $\sigma \in \Sigma_N^Y$, we show in Proposition 4.2 that the performance criterion $\mathbf{E}[g(X_\sigma)|\Pi_0 = \pi]$ can be expressed in terms of the discrete-time Markov chain $(\Pi_n, S_n)_{0 \leq n \leq N}$. We would like to emphasize the following important fact. Although the performance criterion can be written in terms of *discrete-time* process, the optimization problem remains intrinsically a *continuous-time* optimization problem. Indeed, the performance criterion is maximized over the set of stopping times with respect to the *continuous-time* filtration $(\mathfrak{F}_t^Y)_{t \geq 0}$.

Proposition 4.2. *Let $\sigma \in \Sigma^Y$ and $n \geq 1$. For all $\pi \in \mathcal{M}_1(E_0)$ one has*

$$\begin{aligned} & \mathbf{E}[g(X_{\sigma \wedge T_n})|\Pi_0 = \pi] \\ &= \sum_{k=0}^{n-1} \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_k \leq \sigma\}} \mathbb{1}_{\{R_k < t_i^*\}} g \circ \Phi(x_i, R_k) e^{-\Lambda(x_i, R_k)} \Pi_k^i | \Pi_0 = \pi] \\ & \quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_n \leq \sigma\}} g(x_i) \Pi_n^i | \Pi_0 = \pi], \end{aligned}$$

where $(R_k)_{k \in \mathbb{N}}$ is the sequence of non negative random variables associated to σ as introduced in Theorem A.5.

Proof We split $\mathbf{E}[g(X_{\sigma \wedge T_n})|\Pi_0 = \pi]$ into several terms depending on the position of σ w.r.t. the jump times T_k

$$\begin{aligned} \mathbf{E}[g(X_{\sigma \wedge T_n}) | \Pi_0 = \pi] &= \sum_{k=0}^{n-1} \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_k \leq \sigma < T_{k+1}\}} \mathbb{1}_{\{Z_k = x_i\}} g \circ \Phi(x_i, R_k) | \Pi_0 = \pi] \\ & \quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_n \leq \sigma\}} \mathbb{1}_{\{Z_n = x_i\}} g(x_i) | \Pi_0 = \pi]. \end{aligned}$$

For notational convenience, consider

$$\begin{cases} A_{k,i} &= \mathbb{1}_{\{T_k \leq \sigma < T_{k+1}\}} \mathbb{1}_{\{Z_k = x_i\}} g \circ \Phi(x_i, R_k), \\ B_i &= \mathbb{1}_{\{T_n \leq \sigma\}} \mathbb{1}_{\{Z_n = x_i\}} g(x_i). \end{cases}$$

On the one hand, one has $\mathbf{E}[B_i | \mathfrak{F}_{T_n}^Y] = g(x_i) \mathbb{1}_{\{T_n \leq \sigma\}} \Pi_n^i$ since $\{T_n \leq \sigma\} \in \mathfrak{F}_{T_n}^Y$ (see for instance [6, p. 298, Theorem T7]). On the other hand, to compute $\mathbf{E}[A_{k,i} | \mathfrak{F}_{T_k}^Y]$, we use Lemma A.6 to obtain

$$\begin{aligned} \mathbf{E}[A_{k,i} | \mathfrak{F}_{T_k}^Y] &= \mathbb{1}_{\{T_k \leq \sigma\}} g \circ \Phi(x_i, R_k) \mathbf{E}[\mathbb{1}_{\{S_{k+1} > R_k\}} \mathbb{1}_{\{Z_k = x_i\}} | \mathfrak{F}_{T_k}^Y] \\ &= \mathbb{1}_{\{T_k \leq \sigma\}} g \circ \Phi(x_i, R_k) \mathbf{E}[\mathbb{1}_{\{Z_k = x_i\}} \mathbf{E}[\mathbb{1}_{\{S_{k+1} > R_k\}} | \mathfrak{F}_{T_k}] | \mathfrak{F}_{T_k}^Y] \\ &= \mathbb{1}_{\{T_k \leq \sigma\}} g \circ \Phi(x_i, R_k) \mathbf{E}[\mathbb{1}_{\{Z_k = x_i\}} \mathbb{1}_{\{R_k < t_i^*(Z_k)\}} e^{-\Lambda(Z_k, R_k)} | \mathfrak{F}_{T_k}^Y] \\ &= \mathbb{1}_{\{T_k \leq \sigma\}} g \circ \Phi(x_i, R_k) \mathbb{1}_{\{R_k < t_i^*\}} e^{-\Lambda(x_i, R_k)} \Pi_k^i. \end{aligned}$$

Details to obtain the third line in the above computations are provided by Lemma B.1. The result follows. \square

4.2 Dynamic programming equation

Based on the new formulation, the main objective of this section is to derive the backward dynamic programming equation. It involves some operators introduced in Definition 4.3. By iterating the operator, labelled L , we define a sequence of real valued functions $(v_n)_{0 \leq n \leq N}$ in Definition 4.4. Theorem 4.5 establishes that v_n is the value function of our partially observed optimal stopping problem with horizon T_{N-n} and in particular that v_0 is the value function of problem defined in equation (3).

Another important result of this section is given by Theorem 4.9 which constructs a sequence of ϵ -optimal stopping times.

Definition 4.3. *The operators $G : B(\mathcal{M}_1(E_0)) \rightarrow B(\mathcal{M}_1(E_0) \times \mathbb{R}^+)$, $H : B(E) \rightarrow B(\mathcal{M}_1(E_0) \times \mathbb{R}^+)$, $J : B(\mathcal{M}_1(E_0)) \times B(E) \rightarrow B(\mathcal{M}_1(E_0) \times \mathbb{R}^+)$, and $L : B(\mathcal{M}_1(E_0)) \times B(E) \rightarrow B(\mathcal{M}_1(E_0))$ are defined for all $(v, h) \in B(\mathcal{M}_1(E_0)) \times B(E)$ and $(\pi, u) \in \mathcal{M}_1(E_0) \times \mathbb{R}^+$ by*

$$\begin{aligned} Gv(\pi, u) &= \mathbf{E}[v(\Pi_1) \mathbb{1}_{\{S_1 \leq u\}} | \Pi_0 = \pi], \\ Hh(\pi, u) &= \mathbf{E}\left[\sum_{i=1}^q h \circ \Phi(x_i, u) \Pi_0^i \mathbb{1}_{\{u < t_i^*\}} \mathbb{1}_{\{S_1 > u\}} | \Pi_0 = \pi\right], \\ J(v, h)(\pi, u) &= Hh(\pi, u) + Gv(\pi, u), \\ L(v, h)(\pi) &= \sup_{u \geq 0} J(v, h)(\pi, u). \end{aligned}$$

Definition 4.4. *The sequence $(v_n)_{0 \leq n \leq N}$ of real-valued functions is defined on $\mathcal{M}_1(E_0)$ by*

$$\begin{cases} v_N(\pi) &= \sum_{i=1}^q g(x_i) \pi^i, \\ v_{n-1}(\pi) &= L(v_n, g)(\pi), \quad 1 \leq n \leq N. \end{cases}$$

The following Theorem is the main result of this section showing that the operator L is the dynamic programming operator associated to the initial optimization problem.

Theorem 4.5. *For all $1 \leq n \leq N$ and $\pi \in \mathcal{M}_1(E_0)$, one has*

$$\sup_{\sigma \in \Sigma_n^Y} \mathbf{E}[g(X_\sigma) | \Pi_0 = \pi] = v_{N-n}(\pi).$$

Proof The proof of this result is based on Proposition 4.6 and Theorem 4.9. Proposition 4.6 proves that v_{N-n} is an upper bound for the value function of the problem with horizon T_n . The reverse inequality is derived in Theorem 4.9 by constructing a sequence of ϵ -optimal stopping times. \square

Proposition 4.6. *For all $1 \leq n \leq N$ and $\pi \in \mathcal{M}_1(E_0)$, one has*

$$\sup_{\sigma \in \Sigma_n^Y} \mathbf{E}[g(X_\sigma) | \Pi_0 = \pi] \leq v_{N-n}(\pi).$$

Proof Let $\sigma \in \Sigma^Y$. Consider $(R_k)_{k \in \mathbb{N}}$ the sequence associated to σ as introduced in Theorem A.5. We prove the theorem by induction on n . For $n = 1$, Proposition 4.2 yields

$$\begin{aligned} \mathbf{E}[g(X_{\sigma \wedge T_1}) | \Pi_0 = \pi] &= \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{R_0 < t_i^*\}} g \circ \Phi(x_i, R_0) e^{-\Lambda(x_i, R_0)} \Pi_0^i | \Pi_0 = \pi] \\ &\quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_1 \leq \sigma\}} g(x_i) \Pi_1^i | \Pi_0 = \pi]. \end{aligned} \quad (8)$$

Since R_0 is deterministic and by using Lemma C.1, we recognize that the first term of the right hand side of equation (8) is $Hg(\pi, R_0)$. We now turn to the second term of the right hand side of equation (8) which is given by

$$\begin{aligned} \mathbf{E}[\mathbb{1}_{\{S_1 \leq R_0\}} \sum_{i=1}^q g(x_i) \Pi_1^i | \Pi_0 = \pi] &= \mathbf{E}[v_N(\Pi_1) \mathbb{1}_{\{S_1 \leq R_0\}} | \Pi_0 = \pi] \\ &= Gv_N(\pi, R_0), \end{aligned}$$

from Lemma A.6 and the definition of G . Recall that from Definition 4.3 one has $J(v_N, g) = Hg + Gv_N$ thus, one obtains

$$\begin{aligned} \mathbf{E}[g(X_{\sigma \wedge T_1}) | \Pi_0 = \pi] &= J(v_N, g)(\pi, R_0) \leq \sup_{u \geq 0} J(v_N, g)(\pi, u) \\ &= L(v_N, g)(\pi) = v_{N-1}(\pi). \end{aligned}$$

Set now $2 \leq n \leq N$ and assume that $\mathbf{E}[g(X_\tau) | \Pi_0 = \pi] \leq v_{N-(n-1)}(\pi)$, for all $\tau \in \Sigma_{n-1}^Y$. Proposition 4.2 yields

$$\begin{aligned} &\mathbf{E}[g(X_{\sigma \wedge T_n}) | \Pi_0 = \pi] \\ &= \sum_{k=0}^{n-1} \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_k \leq \sigma\}} \mathbb{1}_{\{R_k < t_i^*\}} g \circ \Phi(x_i, R_k) e^{-\Lambda(x_i, R_k)} \Pi_k^i | \Pi_0 = \pi] \\ &\quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_n \leq \sigma\}} g(x_i) \Pi_n^i | \Pi_0 = \pi]. \end{aligned}$$

As in the case $n = 1$, the term for $k = 0$ equals $Hg(\pi, R_0)$. Notice that for $k \geq 1$, $\mathbb{1}_{\{T_k \leq \sigma\}} = \mathbb{1}_{\{T_k \leq \sigma\}} \mathbb{1}_{\{T_1 \leq \sigma\}}$ and that $\mathbb{1}_{\{T_1 \leq \sigma\}} = \mathbb{1}_{\{S_1 \leq R_0\}}$ is $\mathfrak{F}_{T_1}^Y$ -measurable. By taking the conditional expectation w.r.t. $\mathfrak{F}_{T_1}^Y$ it follows that $\mathbf{E}[\Xi \mathbb{1}_{\{S_1 \leq R_0\}} | \Pi_0 = \pi] = \mathbf{E}[\Xi | \Pi_0 = \pi]$ where Ξ is defined by

$$\begin{aligned} \Xi &= \mathbf{E} \left[\sum_{k=1}^{n-1} \sum_{i=1}^q \mathbb{1}_{\{T_k \leq \sigma\}} \mathbb{1}_{\{R_k < t_i^*\}} g \circ \Phi(x_i, R_k) e^{-\Lambda(x_i, R_k)} \Pi_k^i \right. \\ &\quad \left. + \sum_{i=1}^q \mathbb{1}_{\{T_n \leq \sigma\}} g(x_i) \Pi_n^i \middle| \mathfrak{F}_{T_1}^Y \right]. \end{aligned}$$

Therefore, we obtain

$$\mathbf{E}[g(X_{\sigma \wedge T_n}) | \Pi_0 = \pi] = Hg(\pi, R_0) + \mathbf{E}[\Xi \mathbb{1}_{\{S_1 \leq R_0\}} | \Pi_0 = \pi]. \quad (9)$$

We now use the Markov property of the chain $(\Pi_k)_{k \geq 0}$. Indeed, for $k \geq 1$, one has $\Pi_k = \Pi_{k-1} \circ \theta$, where θ is the translation operator of the $(\mathfrak{F}_{T_n}^Y)_{n \in \mathbb{N}}$ -Markov chain $(\Pi_n, Y_n, S_n)_{n \in \mathbb{N}}$. Moreover, when $T_1 \leq \sigma$, one has, from Proposition A.10, $R_k = \tilde{R}_{k-1}^1 \circ \theta$ (indeed, we pointed out in Remark A.8 that R_k can be replaced by \bar{R}_k defined in Lemma A.7) and $\sigma = T_1 + \tilde{\sigma} \circ \theta$ where \tilde{R}_{k-1}^1 and $\tilde{\sigma}$ are defined in Definition A.9 and Proposition A.10 (with $l = 1$ in the present case). Since for $k \geq 1$, $T_k = T_1 + T_{k-1} \circ \theta$, one has $\mathbb{1}_{\{T_k \leq \sigma\}} = \mathbb{1}_{\{T_{k-1} \leq \tilde{\sigma}\}} \circ \theta$. Finally, combining the Markov property of the chain $(\Pi_k)_{k \geq 0}$ and Proposition 4.2 we have $\Xi = w(\Pi_1)$ with $w(\pi) = \mathbf{E}[g(X_{\tilde{\sigma} \wedge T_{n-1}}) | \Pi_0 = \pi]$. Moreover, one has $w(\pi) \leq v_{N-(n-1)}(\pi)$ from the induction assumption since $\tilde{\sigma} \wedge T_{n-1} \in \Sigma_{n-1}^Y$ (indeed, both $\tilde{\sigma}$ and T_{n-1} are $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times from Corollary A.12 and Lemma A.1 respectively). One has then

$$\Xi \leq v_{N-(n-1)}(\Pi_1). \quad (10)$$

Finally, combining Eq. (9) and (10), one has

$$\mathbf{E}[g(X_{\sigma \wedge T_n}) | \Pi_0 = \pi] \leq Hg(\pi, R_0) + \mathbf{E}[v_{N-(n-1)}(\Pi_1) \mathbb{1}_{\{S_1 \leq R_0\}} | \Pi_0 = \pi].$$

In the second term, we recognize the operator G and one has

$$\begin{aligned} \mathbf{E}[g(X_{\sigma \wedge T_n}) | \Pi_0 = \pi] &\leq Hg(\pi, R_0) + Gv_{N-(n-1)}(\pi, R_0) \\ &= J(v_{N-(n-1)}, g)(\pi, R_0) \\ &\leq \sup_{u \geq 0} J(v_{N-(n-1)}, g)(\pi, u) \\ &= L(v_{N-(n-1)}, g)(\pi) = v_{N-n}(\pi), \end{aligned}$$

that proves the induction. \square

We now prove the reverse inequality by constructing a sequence of ϵ -optimal stopping times.

Definition 4.7. For $\epsilon > 0$, $1 \leq n \leq N$ and for $\pi \in \mathcal{M}_1(E_0)$, we define

$$r_n^\epsilon(\pi) = \inf \{u > 0 : J(v_{N-n}, g)(\pi, u) > v_{N-n-1}(\pi) - \epsilon\}.$$

Consider $R_{1,0}^\epsilon = r_0^\epsilon(\Pi_0)$ and for $2 \leq n \leq N$,

$$\begin{cases} R_{n,0}^\epsilon &= r_{n-1}^{\epsilon/2}(\Pi_0), \\ R_{n,k}^\epsilon &= r_{n-1-k}^{\epsilon/(2^{k+1})}(\Pi_k) \mathbb{1}_{\{R_{n,k-1}^\epsilon \geq S_k\}} \text{ for } 1 \leq k \leq n-2, \\ R_{n,n-1}^\epsilon &= r_0^{\epsilon/(2^{n-1})}(\Pi_{n-1}) \mathbb{1}_{\{R_{n,n-2}^\epsilon \geq S_{n-1}\}}, \end{cases}$$

and finally set

$$U_n^\epsilon = \sum_{k=1}^n R_{n,k-1}^\epsilon \wedge S_k.$$

The following lemma describes the effect of the translation operator θ on the sequence $(R_{n,k}^\epsilon)_{1 \leq n \leq N, 0 \leq k \leq n-1}$.

Lemma 4.8. For $n \geq 2$ and $1 \leq k \leq n-1$, on the set $\{T_1 \leq U_n^{2\epsilon}\}$, one has

$$R_{n-1,k-1}^\epsilon \circ \theta = R_{n,k}^{2\epsilon}.$$

Proof For $n = 2$, one just has to prove that on the event $\{T_1 \leq U_2^{2\epsilon}\}$, one has $R_{1,0}^\epsilon \circ \theta = R_{2,1}^{2\epsilon}$. Yet, from the definition of the sequence $(R_{n,k}^\epsilon)_{1 \leq n \leq N, 0 \leq k \leq n-1}$, one has $R_{1,0}^\epsilon \circ \theta = r_0^\epsilon(\Pi_1)$ and $R_{2,1}^{2\epsilon} = r_0^{\frac{2\epsilon}{2}}(\Pi_1) \mathbb{1}_{\{R_{2,0}^{2\epsilon} \geq S_1\}}$. The result follows since we are on the event $\{T_1 \leq U_2^{2\epsilon}\} = \{R_{2,0}^{2\epsilon} \geq S_1\}$. For a fixed $n \geq 3$, we prove the lemma by induction on $1 \leq k \leq n-1$. Set $k = 1$. One has from the definition on the sequence $(R_{n,k}^\epsilon)_{1 \leq n \leq N, 0 \leq k \leq n-1}$, $R_{n-1,0}^\epsilon \circ \theta = r_{n-2}^{\frac{\epsilon}{2}}(\Pi_1)$ and $R_{n,1}^{2\epsilon} = r_{n-2}^{\frac{2\epsilon}{4}}(\Pi_1) \mathbb{1}_{\{R_{n,0}^{2\epsilon} \geq S_1\}}$. We obtain $R_{n-1,0}^\epsilon \circ \theta = R_{n,1}^{2\epsilon}$ because we have assumed that we are on the event $\{T_1 \leq U_n^{2\epsilon}\} = \{R_{n,0}^{2\epsilon} \geq S_1\}$. The propagation of the induction is similar to the case $k = 1$. \square

Equipped with this preliminary result, we may now prove that $(U_n^\epsilon)_{1 \leq n \leq N}$ is a sequence of ϵ -optimal stopping times with respect to the filtration. generated by the observations.

Theorem 4.9. *For all $1 \leq n \leq N$ and $\epsilon > 0$, one has $U_n^\epsilon \in \Sigma_n^Y$ and*

$$\mathbf{E}[g(X_{U_n^\epsilon}) | \Pi_0 = \pi] \geq v_{N-n}(\pi) - \epsilon.$$

Proof Let $n \in \{1, \dots, N\}$. First notice that, as a direct consequence of Proposition A.11, U_n^ϵ is an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time since, by construction, the $R_{n,k}^\epsilon$ are $\mathfrak{F}_{T_k}^Y$ -measurable and satisfy the condition $R_{n,k}^\epsilon = 0$ on the event $\{S_k > R_{n,k-1}^\epsilon\}$. It is also clear that $U_n^\epsilon \leq \sum_{k=1}^n S_k = T_n$. Thus, one has $U_n^\epsilon \in \Sigma_n^Y$. Let us now prove the second assessment by induction. Set $n = 1$. Let $\pi \in \mathcal{M}_1(E_0)$, we denote $r_0^\epsilon = r_0^\epsilon(\pi)$. Since $R_{1,0}^\epsilon = r_0^\epsilon$ is deterministic, one has clearly $R_{1,0}^\epsilon \in \Sigma^Y$. Consequently, by using the same arguments as in the proof of Proposition 4.6, we obtain

$$\mathbf{E}[g(X_{R_{1,0}^\epsilon \wedge S_1}) | \Pi_0 = \pi] = Hg(\pi, r_0^\epsilon) + Gv_N(\pi, r_0^\epsilon) = J(v_N, g)(\pi, r_0^\epsilon).$$

Finally, the definition of r_0^ϵ yields $J(v_N, g)(\pi, r_0^\epsilon) \geq v_{N-1}(\pi) - \epsilon$ thus one has

$$\mathbf{E}[g(X_{R_{1,0}^\epsilon \wedge S_1}) | \Pi_0 = \pi] \geq v_{N-1}(\pi) - \epsilon.$$

Now set $2 \leq n \leq N$ and assume that $\mathbf{E}[g(X_{U_{n-1}^\epsilon}) | \Pi_0 = \pi] \geq v_{N-(n-1)}(\pi) - \epsilon$, for all $\epsilon > 0$. Proposition 4.2 yields

$$\begin{aligned} & \mathbf{E}[g(X_{U_n^{2\epsilon}}) | \Pi_0 = \pi] \\ &= \sum_{k=0}^{n-1} \sum_{i=1}^q \mathbf{E} \left[\mathbb{1}_{\{T_k \leq U_n^{2\epsilon}\}} \mathbb{1}_{\{R_{n,k}^{2\epsilon} < t_i^*\}} g \circ \Phi(x_i, R_{n,k}^{2\epsilon}) e^{-\Lambda(x_i, R_{n,k}^{2\epsilon})} \Pi_k^i \middle| \Pi_0 = \pi \right] \\ & \quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_n \leq U_n^{2\epsilon}\}} g(x_i) \Pi_n^i | \Pi_0 = \pi]. \end{aligned}$$

Denote $r_{n-1}^\epsilon = r_{n-1}^\epsilon(\pi)$. As in the case $n = 1$, the term for $k = 0$ equals $Hg(\pi, r_{n-1}^\epsilon)$ since $R_{n,0}^{2\epsilon} = r_{n-1}^\epsilon(\Pi_0)$. Take the conditional expectation w.r.t. $\mathfrak{F}_{T_1}^Y$ in the other terms. One has then,

$$\mathbf{E}[g(X_{U_n^{2\epsilon}}) | \Pi_0 = \pi] = Hg(\pi, r_{n-1}^\epsilon) + \mathbf{E}[\Xi' \mathbb{1}_{\{T_1 \leq U_n^{2\epsilon}\}} | \Pi_0 = \pi], \quad (11)$$

with

$$\begin{aligned} \Xi' &= \mathbf{E} \left[\sum_{k=1}^{n-1} \sum_{i=1}^q \mathbb{1}_{\{T_k \leq U_n^{2\epsilon}\}} \mathbb{1}_{\{R_{n,k}^{2\epsilon} < t_i^*\}} g \circ \Phi(x_i, R_{n,k}^{2\epsilon}) e^{-\Lambda(x_i, R_{n,k}^{2\epsilon})} \Pi_k^i \right. \\ &\quad \left. + \sum_{i=1}^q \mathbb{1}_{\{T_n \leq U_n^{2\epsilon}\}} g(x_i) \Pi_n^i \middle| \mathfrak{F}_{T_1}^Y \right]. \end{aligned}$$

Our objective is to apply the Markov property of $(\Pi_k)_{k \in \mathbb{N}}$ in the term Ξ' . Recall that, from Lemma 4.8, one has $R_{n-1,k-1}^\epsilon \circ \theta = R_{n,k}^{2\epsilon}$ for $n \geq 2$ and $1 \leq k \leq n-1$ on the event $\{T_1 \leq U_n^{2\epsilon}\} = \{S_1 \leq R_{n,0}^{2\epsilon}\}$ (the equality of these events stems from Lemma A.6). Thus, on this set one has

$$\begin{aligned} U_n^{2\epsilon} &= S_1 + \sum_{k=2}^n R_{n,k-1}^{2\epsilon} \wedge S_k = T_1 + \sum_{k=2}^n (R_{n-1,k-2}^\epsilon \circ \theta) \wedge (S_{k-1} \circ \theta) \\ &= T_1 + U_{n-1}^\epsilon \circ \theta. \end{aligned}$$

Besides, recall that $T_k = T_1 + T_{k-1} \circ \theta$, for $k \geq 1$. Consequently, on the set $\{T_1 \leq U_n^{2\epsilon}\}$, one has $\mathbb{1}_{\{T_k \leq U_n^{2\epsilon}\}} = \mathbb{1}_{\{T_{k-1} \leq U_{n-1}^\epsilon\}} \circ \theta$ and thus, combining the Markov property of the chain $(\Pi_k)_{k \geq 0}$ and Proposition 4.2, we have

$$\Xi'(\Pi_1) = w'(\Pi_1),$$

with $w'(\pi) = \mathbf{E} \left[g(X_{U_{n-1}^\epsilon}) \middle| \Pi_0 = \pi \right]$. Moreover, thanks to the induction assumption, one has $w'(\pi) \geq v_{N-(n-1)}(\pi) - \epsilon$ so that one obtains

$$\Xi' \geq v_{N-(n-1)}(\Pi_1) - \epsilon. \quad (12)$$

Finally, combining equation (11) and (12) and noticing that, according to Lemma A.6, $\{T_1 \leq U_n^{2\epsilon}\} = \{S_1 \leq r_{n-1}^\epsilon\}$, one obtains

$$\begin{aligned} \mathbf{E}[g(X_{U_n^{2\epsilon}}) | \Pi_0 = \pi] &\geq Hg(\pi, r_{n-1}^\epsilon) + \mathbf{E}[v_{N-(n-1)}(\Pi_1) \mathbb{1}_{\{S_1 \leq r_{n-1}^\epsilon\}} | \Pi_0 = \pi] - \epsilon \\ &= J(v_{N-(n-1)}, g)(\pi, r_{n-1}^\epsilon) - \epsilon \\ &\geq v_{N-n}(\pi) - 2\epsilon, \end{aligned}$$

from the definition of r_{n-1}^ϵ , showing the result. \square

5 Numerical approximation by quantization

In this section, we are interested in the computational issue for our optimal stopping problem under partial observation. Indeed, we want to compute a numerical approximation of the value function (3) and propose a computable ϵ -optimal stopping time.

As we have seen in the previous section, the value function v can be obtained by iterating the dynamic programming operator L . However, the operator L involves conditional expectations that are in essence difficult to compute and iterate numerically. We manage to overcome this difficulty by combining two special properties

of our problem. On the one hand, the underlying process (Π_n, S_n) in the expression of the operator L is a Markov chain. Therefore, it can be discretized using a quantization technique which is a powerful method suitable for numerical computation and iteration of conditional expectations. On the other hand, the recursion on the functions $(v_n)_{0 \leq n \leq N}$ involving the operator L can be transformed into a recursion on suitably defined random variables. Thus they are easier to iterate numerically as we do not need to compute an approximation of each v_n on the whole state space.

This section is organized as follows. We first explain how the recursion on the functions $(v_n)_{0 \leq n \leq N}$ can be transformed into a recurrence on random variables involving only the Markov chain (Π_n, S_n) . Then, we present a quantization technique to discretize this Markov chain. Afterwards, we construct a discretized version of the main operators in Definition 5.6 that is used to build an approximation of the value function in Definition 5.7, and a computable ϵ -optimal stopping time. The main results of this section are Theorems 5.8 and 5.17 that prove the convergence of our approximation scheme and provide a rate of convergence.

We first explain how the dynamic programming equations on the functions $(v_n)_{0 \leq n \leq N}$ yield a recursion on the random variables $(v_n(\Pi_n))_{0 \leq n \leq N}$. Introduce now the sequence $(V_n)_{0 \leq n \leq N}$ of random variables defined by

$$V_n = v_n(\Pi_n).$$

In other words, one has

$$\begin{aligned} V_N &= \sum_{i=1}^q g(x_i) \Pi_N^i, \\ V_n &= \sup_{u \geq 0} \mathbf{E} \left[\sum_{i=1}^q g \circ \Phi(x_i, u) \Pi_n^i \mathbb{1}_{\{u < t_i^*\}} \mathbb{1}_{\{S_{n+1} > u\}} + V_{n+1} \mathbb{1}_{\{S_{n+1} \leq u\}} | \Pi_n \right], \end{aligned} \tag{13}$$

for $0 \leq n \leq N - 1$. Notice that V_N is known and the expression of V_n involves only V_{n+1} and the Markov chain (Π_n, S_n) . Thus, the sequence $(V_n)_{0 \leq n \leq N}$ is completely characterized by the system (13). In addition, $V_0 = v_0(\Pi_0) = v(\Pi_0)$. Thus to approximate the value function v at the initial point of our process, it is sufficient to provide an approximation of the sequence of random variables $(V_n)_{0 \leq n \leq N}$.

5.1 The quantization approach

There exists an extensive literature on quantization methods for random variables and processes. We do not pretend to present here an exhaustive panorama of these methods. However, the interested reader may for instance, consult the following works [2, 12, 16] and references therein. Consider X an \mathbb{R}^r -valued random variable such that $\|X\|_p < \infty$ where $\|X\|_p$ denotes the L^p -nom of X : $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$. Let ν be a fixed integer, the optimal L^p -quantization of the random variable X consists in finding the best possible L^p -approximation of X by a random vector \widehat{X} taking at most ν values: $\widehat{X} \in \{x^1, \dots, x^\nu\}$. This procedure consists in the following two steps:

1. Find a finite weighted grid $\Gamma \subset \mathbb{R}^r$ with $\Gamma = \{x^1, \dots, x^\nu\}$.

2. Set $\widehat{X} = \widehat{X}^\Gamma$ where $\widehat{X}^\Gamma = \text{proj}_\Gamma(X)$ with proj_Γ denotes the closest neighbour projection on Γ .

The asymptotic properties of the L^p -quantization are given by the following result, see e.g. [16].

Theorem 5.1. *If $\mathbb{E}[|X|^{p+\eta}] < +\infty$ for some $\eta > 0$ then one has*

$$\lim_{\nu \rightarrow \infty} \nu^{p/r} \min_{|\Gamma| \leq \nu} \|X - \widehat{X}^\Gamma\|_p^p = J_{p,r} \left(\int |h|^{r/(r+p)}(u) du \right)^{1+p/r},$$

where the distribution of X is $P_X(du) = h(u)\lambda_r(du) + \mu$ with $\mu \perp \lambda_r$, $J_{p,r}$ a constant and λ_r the Lebesgue measure in \mathbb{R}^r .

There exists a similar procedure for the optimal quantization of a Markov chain. Our approximation method is based on the quantization of the Markov chain $(\Pi_k, S_k)_{k \leq N}$. Thus, from now on, we will denote, for $0 \leq k \leq N$, $\Theta_k = (\Pi_k, S_k)$. The CLVQ (Competitive Learning Vector Quantization) algorithm [2, Section 3] provides for each time step $0 \leq k \leq N$ a finite grid Γ_k of $\mathcal{M}_1(E_0) \times \mathbb{R}^+$ as well as the transition matrices $(\widehat{Q}_k)_{0 \leq k \leq N-1}$ from Γ_k to Γ_{k+1} . Let $p \geq 1$ such that for all $k \leq N$, Π_k and S_k have finite moments at least up to order p and let proj_{Γ_k} be the nearest-neighbor projection from $\mathcal{M}_1(E_0) \times \mathbb{R}^+$ onto Γ_k . The quantized process $(\widehat{\Theta}_k)_{k \leq N} = (\widehat{\Pi}_k, \widehat{S}_k)_{k \leq N}$ with value for each k in the finite grid Γ_k of $\mathcal{M}_1(E_0) \times \mathbb{R}^+$ is then defined by

$$(\widehat{\Pi}_k, \widehat{S}_k) = \text{proj}_{\Gamma_k}(\Pi_k, S_k).$$

We will also denote by Γ_k^Π , the projection of Γ_k on $\mathcal{M}_1(E_0)$, and by Γ_k^S , the projection of Γ_k on \mathbb{R}^+ .

Some important remarks must be made concerning the quantization. On the one hand, the optimal quantization has nice convergence properties stated by Theorem 5.1. Indeed, the L^p -quantization error $\|\Theta_k - \widehat{\Theta}_k\|_p$ goes to zero when the number of points in the grids goes to infinity. However, on the other hand, the Markov property is not maintained by the algorithm and the quantized process is generally not Markovian. Although the quantized process can be easily transformed into a Markov chain, this chain will not be homogeneous. It must be pointed out that the quantized process $(\widehat{\Theta}_k)_{k \in \mathbb{N}}$ depends on the starting point Θ_0 of the process.

In practice, we begin with the computation of the quantization grids, which merely requires to be able to simulate the process. Notice that in our case, what is actually simulated is the sequence of observation $(Y_k, S_k)_{0 \leq k \leq N}$. We are then able to compute the filter $(\Pi_k)_{0 \leq k \leq N}$ thanks to the recursive equation provided by Proposition 3.4. The grids are only computed once and for all and may be stored off-line. Our schemes are then based on the following simple idea: we replace the process by its quantized approximation within the different recursions. The computation is thus carried out in a very simple way since the quantized process has finite state space.

5.2 Approximation of the value function

Our approximation scheme of the sequence $(V_n)_{0 \leq n \leq N}$ follows the same lines as in [9], but once more, the results therein cannot be applied directly as the Markov

chain $(\Theta_k)_{k \in \mathbb{N}}$ is not the underlying Markov chain of some PDMP. Our approach decomposes in two steps. The first one will be to discretize the time-continuous maximization of the operator L to obtain a maximization over a finite set. The second step consists in replacing the Markov chain $(\Theta_n)_{n \in \mathbb{N}} = (\Pi_n, S_n)_{n \in \mathbb{N}}$ by its quantized approximation $(\widehat{\Theta}_n)_{n \in \mathbb{N}} = (\widehat{\Pi}_n, \widehat{S}_n)_{n \in \mathbb{N}}$ within the dynamic programming equation. Thus, the conditional expectations will become easily tractable finite sums.

Let us first build a finite time grid to discretize the continuous-time maximization in the expression of the operator L . The maximum is originally taken over the set $[0, \infty[$. However, it can be seen from Definition 4.3 that $J(v, h)(\pi, u) = J(v, h)(\pi, t_q^*)$ for all $u \geq t_q^*$. Indeed, the random variable S_1 is bounded by the greatest deterministic exit time t_q^* that is finite thanks to Assumption 2.3. Therefore, the maximization set can be reduced to the compact set $[0, t_q^*]$. Instead of directly discretizing the set $[0, t_q^*]$, we will actually discretize the subsets $]t_m^*, t_{m+1}^*]$. The reason why we want to exclude the points t_m^* from our grid is technical and will be explained with Lemma 5.12. Now, it seems natural to distinguish whether $t_m^* = t_{m+1}^*$ or $t_m^* < t_{m+1}^*$.

Definition 5.2. *Let $M \subset \{0, \dots, q-1\}$ be the set of indices m such that $t_m^* < t_{m+1}^*$.*

Notice that M is not empty because it contains at least the index 0 since we assumed that $t_1^* > 0 = t_0^*$. We can now build our approximation grid.

Definition 5.3. *Let $\Delta > 0$ be such that*

$$\Delta < \frac{1}{2} \min \left\{ |t_i^* - t_j^*| \text{ with } 0 \leq i, j, \leq q \text{ such that } t_i^* \neq t_j^* \right\}. \quad (14)$$

For all $m \in M$, let $Gr_m(\Delta)$ be the finite grid on $]t_m^; t_{m+1}^*]$ defined as follows*

$$Gr_m(\Delta) = \{t_m^* + i\Delta, 1 \leq i \leq i_m\} \cup \{t_{m+1}^* - \Delta\},$$

where $i_m = \max\{i \in \mathbb{N} \text{ such that } t_m^ + i\Delta \leq t_{m+1}^* - \Delta\}$. We also denote $Gr(\Delta) = \cup_{m \in M} Gr_m(\Delta)$.*

Remark 5.4. *Let $m \in M$. Notice that, thanks to Eq. (14), $Gr_m(\Delta)$ is not empty. Moreover, it satisfies two properties that will be crucial in the sequel:*

- a.** *for all $t \in]t_m^*; t_{m+1}^*]$, there exists $u \in Gr_m(\Delta)$ such that $|u - t| \leq \Delta$,*
- b.** *for all $u \in Gr_m(\Delta)$ and $0 < \eta < \Delta$, one has $[u - \eta; u + \eta] \subset]t_m^*; t_{m+1}^*]$.*

A discretized maximization operator L^d is then defined as follows.

Definition 5.5. *Let $L^d: B(\mathcal{M}_1(E_0)) \times B(E) \rightarrow B(\mathcal{M}_1(E_0))$ be defined for all $\pi \in \mathcal{M}_1(E_0)$ by*

$$L^d(v, h)(\pi) = \max_{m \in M} \left\{ \max_{u \in Gr_m(\Delta)} \{J(v, h)(\pi, u)\} \right\} \vee Kv(\pi),$$

with $Kv(\pi) = J(v, h)(\pi, t_q^) = Gv(\pi, t_q^*) = \mathbf{E}[v(\Pi_1) | \Pi_0 = \pi]$.*

We now proceed to our second step: replacing the Markov chain $(\Theta_n)_{n \in \mathbb{N}} = (\Pi_n, S_n)_{n \in \mathbb{N}}$ by its quantized approximation $(\hat{\Theta}_n)_{n \in \mathbb{N}} = (\hat{\Pi}_n, \hat{S}_n)_{n \in \mathbb{N}}$ within the operators involved in the construction of the value function.

Definition 5.6. We define the quantized operators $\hat{G}_n, \hat{H}_n, \hat{J}_n, \hat{K}_n$ and \hat{L}_n^d for $n \in \{1, \dots, N\}$, $v \in B(\Gamma_n)$, $h \in B(E)$, $\pi \in \Gamma_{n-1}^\Pi$ and $u \geq 0$ as follows

$$\begin{aligned}\hat{G}_n v(\pi, u) &= \mathbf{E}[v(\hat{\Pi}_n) \mathbb{1}_{\{\hat{S}_n \leq u\}} | \hat{\Pi}_{n-1} = \pi], \\ \hat{H}_n h(\pi, u) &= \sum_{i=1}^q \pi^i \mathbb{1}_{\{u < t_i^*\}} h \circ \Phi(x_i, u) \mathbf{E}[\mathbb{1}_{\{\hat{S}_n > u\}} | \hat{\Pi}_{n-1} = \pi], \\ \hat{J}_n(v, h)(\pi, u) &= \hat{H}_n h(\pi, u) + \hat{G}_n v(\pi, u), \\ \hat{K}_n v(\pi) &= \hat{J}_n(v, h)(\pi, t_q^*) = \mathbf{E}[v(\hat{\Pi}_n) | \hat{\Pi}_{n-1} = \pi], \\ \hat{L}_n^d(v, h)(\pi) &= \max_{m \in M} \left\{ \max_{u \in Gr_m(\Delta)} \{ \hat{J}_n(v, h)(\pi, u) \} \right\} \vee \hat{K}_n v(\pi).\end{aligned}$$

The quantized approximation of the value functions naturally follows.

Definition 5.7. For $0 \leq n \leq N$, define the functions \hat{v}_n on Γ_n^Π as follows

$$\begin{cases} \hat{v}_N(\pi) = \sum_{i=1}^q g(x_i) \pi^i & \text{for all } \pi \in \Gamma_N^\Pi, \\ \hat{v}_{n-1}(\pi) = \hat{L}_n^d(\hat{v}_n, g)(\pi) & \text{for all } \pi \in \Gamma_{n-1}^\Pi \text{ and } 1 \leq n \leq N. \end{cases}$$

For $0 \leq n \leq N$, let $\hat{V}_n = \hat{v}_n(\hat{\Pi}_n)$.

We may now state our main result for the numerical approximation.

Theorem 5.8. Suppose that for all $0 \leq n \leq N-1$,

$$\Delta > (2C_\lambda)^{-1/2} \|S_{n+1} - \hat{S}_{n+1}\|_p^{1/2}, \quad (15)$$

then, one has the following bound for the approximation error

$$\begin{aligned}\|V_n - \hat{V}_n\|_p &\leq \|V_{n+1} - \hat{V}_{n+1}\|_p + a\Delta + b\|S_{n+1} - \hat{S}_{n+1}\|_p^{1/2} \\ &\quad + c_n \|\Pi_n - \hat{\Pi}_n\|_p + 2[v_{n+1}] \|\Pi_{n+1} - \hat{\Pi}_{n+1}\|_p,\end{aligned}$$

where $a = [g]_2 + 2C_g C_\lambda$, $b = 2C_g (2C_\lambda)^{1/2}$ and $c_n = [v_n] + 4C_g + 2[v_{n+1}]$ with $[v_n]$, $[v_{n+1}]$ defined in Proposition C.10 and $[g]_2$ defined in Assumption 2.7.

Theorem 5.8 establishes the convergence of our approximation scheme and provides a bound for the rate of convergence. More precisely, it gives a rate for the L^p convergence of \hat{V}_0 towards V_0 . Indeed, one has $\|V_N - \hat{V}_N\|_p = \|\sum_{i=1}^q g(x_i) (\Pi_N^i - \hat{\Pi}_N^i)\|_p \leq C_g \|\Pi_N - \hat{\Pi}_N\|_p$, so by virtue of Theorem 5.8 $|V_0 - \hat{V}_0|$ can be made arbitrarily small when the quantization errors $(\|\Theta_n - \hat{\Theta}_n\|_p)_{0 \leq n \leq N}$ go to zero i.e. when the number of points in the quantization grids goes to infinity.

In order to prove Theorem 5.8, we proceed similarly to [9] and split the approximation error into four terms $\|V_n - \widehat{V}_n\|_p \leq \Xi_1 + \Xi_2 + \Xi_3 + \Xi_4$, with

$$\begin{aligned}\Xi_1 &= \|v_n(\Pi_n) - v_n(\widehat{\Pi}_n)\|_p, \\ \Xi_2 &= \|L(v_{n+1}, g)(\widehat{\Pi}_n) - L^d(v_{n+1}, g)(\widehat{\Pi}_n)\|_p, \\ \Xi_3 &= \|L^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n)\|_p, \\ \Xi_4 &= \|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{\Pi}_n)\|_p.\end{aligned}$$

To obtain bounds for each of these terms, one needs to study the regularity of the operators and the value functions v_n . The results are detailed in Appendix C. In particular, we establish in Proposition C.10 that the value functions v_n are Lipschitz continuous, yielding a bound for the first term.

Lemma 5.9. *The first term Ξ_1 is bounded as follows*

$$\|v_n(\Pi_n) - v_n(\widehat{\Pi}_n)\|_p \leq [v_n] \|\Pi_n - \widehat{\Pi}_n\|_p.$$

The other error terms are studied separately in the following sections.

5.2.1 Second term of the error

For the second error term, we investigate the consequences of replacing the continuous maximization in operator L by a discrete one on $Gr(\Delta)$.

Lemma 5.10. *For all $m \in M$, $v \in B(\mathcal{M}_1(E_0))$ and $\pi \in \mathcal{M}_1(E_0)$ one has*

$$\left| \sup_{u \in [t_m^*; t_{m+1}^*]} J(v, g)(\pi, u) - \max_{u \in Gr_m(\Delta)} J(v, g)(\pi, u) \right| \leq ([g]_2 + C_g C_\lambda + C_v C_\lambda) \Delta.$$

Proof We use Definition C.2 to split operator J into a sum of continuous operators J^m . Thus, one has

$$\sup_{u \in [t_m^*; t_{m+1}^*]} J(v, g)(\pi, u) = \sup_{u \in [t_m^*; t_{m+1}^*]} J^m(v, g)(\pi, u).$$

The function $u \rightarrow J^m(v, h)(\pi, u)$ being continuous, there exists $\bar{t} \in [t_m^*; t_{m+1}^*]$ such that $\sup_{u \in [t_m^*; t_{m+1}^*]} J^m(v, h)(\pi, u) = J^m(v, h)(\pi, \bar{t})$. Moreover, from Remark 5.4.a, one may chose $\bar{u} \in Gr_m(\Delta)$ so that $|\bar{u} - \bar{t}| \leq \Delta$. Propositions C.4 and C.7 stating the Lipschitz continuity of J^m then yield

$$\begin{aligned}0 &\leq \sup_{u \in [t_m^*; t_{m+1}^*]} J^m(v, h)(\pi, u) - \max_{u \in Gr_m(\Delta)} J^m(v, h)(\pi, u) \\ &\leq J^m(v, h)(\pi, \bar{t}) - J^m(v, h)(\pi, \bar{u}) \\ &\leq ([g]_2 + C_g C_\lambda + C_v C_\lambda) |\bar{t} - \bar{u}| \leq ([g]_2 + C_g C_\lambda + C_v C_\lambda) \Delta,\end{aligned}$$

showing the result. □

Lemma 5.11. *The second term Ξ_2 is bounded as follows*

$$\|L(v_{n+1}, g)(\widehat{\Pi}_n) - L^d(v_{n+1}, g)(\widehat{\Pi}_n)\|_p \leq ([g]_2 + 2C_g C_\lambda) \Delta.$$

Proof This is a straightforward consequence of the previous lemma once it has been noticed that for all $a, b, c, d \in \mathbb{R}$, one has $|a \vee b - c \vee d| \leq |a - c| \vee |b - d|$. Notice also that Proposition C.10 provides $C_{v_{n+1}} \leq C_g$. \square

5.2.2 Third term of the error

To investigate the third error term, we use the properties of quantization to bound the error made by replacing an operator by its quantized approximation. As in [9], we must first deal with non-continuous indicator functions. The fact that the t_m^* and a small neighborhood around them do not belong to the discretization grid $Gr(\Delta)$ is crucial to obtain the following lemma.

Lemma 5.12. *For all $0 \leq n \leq N - 1$, $m \in M$ and $0 < \eta < \Delta$, one has*

$$\left\| \max_{u \in Gr_m(\Delta)} \mathbf{E} \left[\left| \mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}} \right| \widehat{\Pi}_n \right] \right\|_p \leq \eta^{-1} \|S_{n+1} - \widehat{S}_{n+1}\|_p + 2\eta C_\lambda.$$

Proof Let $0 < \eta < \Delta$. The difference of the indicator functions equals 1 if and only if S_{n+1} and \widehat{S}_{n+1} are on different sides of u . Therefore, if the difference of the indicator functions equals 1, either $|S_{n+1} - u| \leq \eta$, or $|S_{n+1} - u| > \eta$ and in the latter case $|S_{n+1} - \widehat{S}_{n+1}| > \eta$ too since $|S_{n+1} - \widehat{S}_{n+1}| > |S_{n+1} - u|$. One has $|\mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}}| \leq \mathbb{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta\}} + \mathbb{1}_{\{|S_{n+1} - u| \leq \eta\}}$, leading to

$$\begin{aligned} \left\| \max_{u \in Gr_m(\Delta)} \mathbf{E} \left[\left| \mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}} \right| \widehat{\Pi}_n \right] \right\|_p \\ \leq \left\| \mathbb{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta\}} \right\|_p + \left\| \max_{u \in Gr_m(\Delta)} \mathbf{E} \left[\mathbb{1}_{\{|S_{n+1} - u| \leq \eta\}} \widehat{\Pi}_n \right] \right\|_p. \end{aligned}$$

On the one hand, Markov inequality yields

$$\left\| \mathbb{1}_{\{|S_{n+1} - \widehat{S}_{n+1}| > \eta\}} \right\|_p = \mathbf{P}(|S_{n+1} - \widehat{S}_{n+1}| > \eta)^{\frac{1}{p}} \leq \|S_{n+1} - \widehat{S}_{n+1}\|_p \eta^{-1}.$$

On the other hand, since $u \in Gr_m(\Delta)$, one has $[u - \eta; u + \eta] \subset]t_m^*; t_{m+1}^*[$ [from Remark 5.4.b, thus S_{n+1} has an absolutely continuous distribution on the interval $[u - \eta; u + \eta]$ since it does not contain any of the t_i^* . Besides, recall that $\widehat{\Theta}_n = \text{proj}_{\Gamma_n}(\Theta_n)$, hence, the following inclusions of σ -fields $\sigma(\widehat{\Pi}_n) \subset \sigma(\widehat{\Theta}_n) \subset \sigma(\Theta_n)$. We also have $\sigma(\Theta_n) \subset \mathfrak{F}_{T_n}^Y \subset \mathfrak{F}_{T_n}$, the law of iterated conditional expectations provides

$$\begin{aligned} \mathbf{E} \left[\mathbb{1}_{\{|S_{n+1} - u| \leq \eta\}} \widehat{\Pi}_n \right] &= \mathbf{E} \left[\mathbf{E} \left[\mathbf{E} \left[\mathbb{1}_{\{|S_{n+1} - u| \leq \eta\}} \middle| \mathfrak{F}_{T_n} \right] \middle| \mathfrak{F}_{T_n}^Y \right] \widehat{\Pi}_n \right] \\ &\leq \mathbf{E} \left[\mathbf{E} \left[\int_{u-\eta}^{u+\eta} \lambda(\Phi(Z_n, s)) ds \middle| \mathfrak{F}_{T_n}^Y \right] \widehat{\Pi}_n \right] \\ &= \mathbf{E} \left[\sum_{i=1}^q \Pi_n^i \int_{u-\eta}^{u+\eta} \lambda(\Phi(x_i, s)) ds \middle| \widehat{\Pi}_n \right]. \end{aligned}$$

Finally, one obtains $\mathbf{E} \left[\mathbb{1}_{\{|S_{n+1} - u| \leq \eta\}} \widehat{\Pi}_n \right] \leq 2\eta C_\lambda$, showing the result. \square

Lemma 5.13. *For all $0 \leq n \leq N - 1$, one has*

$$\begin{aligned} |Kv_{n+1}(\widehat{\Pi}_n) - \widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n)| \\ \leq [v_{n+1}] \mathbf{E} \left[|\Pi_{n+1} - \widehat{\Pi}_{n+1}| \middle| \widehat{\Pi}_n \right] + (2C_g + 2[v_{n+1}]) \mathbf{E} \left[|\Pi_n - \widehat{\Pi}_n| \middle| \widehat{\Pi}_n \right]. \end{aligned}$$

Proof By the definitions of operators K and \widehat{K}_{n+1} , one has

$$\begin{aligned}
& |Kv_{n+1}(\widehat{\Pi}_n) - \widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n)| \\
&= |\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n] - \mathbf{E}[v_{n+1}(\widehat{\Pi}_{n+1})|\widehat{\Pi}_n]| \\
&\leq |\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n] - \mathbf{E}[v_{n+1}(\Pi_{n+1})|\widehat{\Pi}_n]| \\
&\quad + |\mathbf{E}[v_{n+1}(\Pi_{n+1}) - v_{n+1}(\widehat{\Pi}_{n+1})|\widehat{\Pi}_n]|.
\end{aligned} \tag{16}$$

The second term in the right-hand side of Eq. (16) is readily bounded by using Proposition C.10 stating that v_{n+1} is Lipschitz continuous

$$|\mathbf{E}[v_{n+1}(\Pi_{n+1}) - v_{n+1}(\widehat{\Pi}_{n+1})|\widehat{\Pi}_n]| \leq [v_{n+1}]\mathbf{E}[|\Pi_{n+1} - \widehat{\Pi}_{n+1}||\widehat{\Pi}_n].$$

To deal with the first term in the right-hand side of Eq. (16), we need to use the special properties of quantization. Indeed, one has $(\widehat{\Pi}_n, \widehat{S}_n) = \text{proj}_{\Gamma_n}(\Pi_n, S_n)$ so that we have the inclusion of σ -fields $\sigma(\widehat{\Pi}_n) \subset \sigma(\Pi_n, S_n)$. The law of iterated conditional expectations gives

$$\mathbf{E}[v_{n+1}(\Pi_{n+1})|\widehat{\Pi}_n] = \mathbf{E}[\mathbf{E}[v_{n+1}(\Pi_{n+1})|(\Pi_n, S_n)]|\widehat{\Pi}_n].$$

Moreover, Proposition 4.1 yields $\mathbf{E}[v_{n+1}(\Pi_{n+1})|(\Pi_n, S_n)] = \mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n]$, as the conditional distribution of Π_{n+1} w.r.t. (Π_n, S_n) merely depends on Π_n . In addition, $|\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n]|$ is $\sigma(\widehat{\Pi}_n)$ -measurable. One has then

$$\begin{aligned}
& |\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n] - \mathbf{E}[v_{n+1}(\Pi_{n+1})|\widehat{\Pi}_n]| \\
&= |\mathbf{E}[\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n] - \mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n]|\widehat{\Pi}_n]| \\
&= |\mathbf{E}[Kv_{n+1}(\widehat{\Pi}_n) - Kv_{n+1}(\Pi_n)|\widehat{\Pi}_n]|,
\end{aligned}$$

by definition of K . Finally, one has

$$\begin{aligned}
& |\mathbf{E}[v_{n+1}(\Pi_{n+1})|\Pi_n = \widehat{\Pi}_n] - \mathbf{E}[v_{n+1}(\Pi_{n+1})|\widehat{\Pi}_n]| \\
&\leq 2(C_g + [v_{n+1}])\mathbf{E}[|\Pi_n - \widehat{\Pi}_n||\widehat{\Pi}_n],
\end{aligned}$$

thanks to Propositions C.8 and C.10 stating the Lipschitz continuity of operator K and function v_{n+1} . \square

Lemma 5.14. *If Δ satisfies Condition (15), a upper bound for the third term Ξ_3 is*

$$\begin{aligned}
& \|L^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n)\|_p \\
&\leq [v_{n+1}]\|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p + (4C_g + 2[v_{n+1}])\|\Pi_n - \widehat{\Pi}_n\|_p \\
&\quad + 2C_g(2C_\lambda)^{1/2}\|S_{n+1} - \widehat{S}_{n+1}\|_p^{1/2}.
\end{aligned}$$

Proof One has

$$\begin{aligned}
& |L^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n)| \\
&\leq \max_{m \in M} \left\{ \max_{u \in Gr_m(\Delta)} |J(v_{n+1}, g)(\widehat{\Pi}_n, u) - \widehat{J}_{n+1}(v_{n+1}, g)(\widehat{\Pi}_n, u)| \right\} \\
&\quad \vee |Kv_{n+1}(\widehat{\Pi}_n) - \widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n)|.
\end{aligned}$$

The term involving operator K was studied in the previous lemma. Let us now study the term involving operator J . Set m in M , u in $Gr_m(\Delta)$ and define $\alpha(\pi, \pi', s') = \sum_{i=1}^q \pi^i g(\Phi(x_i, u)) \mathbb{1}_{\{s' > u\}} + v_{n+1}(\pi') \mathbb{1}_{\{s' \leq u\}}$. One has then

$$\begin{aligned} & |J(v_{n+1}, g)(\widehat{\Pi}_n, u) - \widehat{J}_{n+1}(v_{n+1}, g)(\widehat{\Pi}_n, u)| \\ &= |J^m(v_{n+1}, g)(\widehat{\Pi}_n, u) - \widehat{J}_{n+1}(v_{n+1}, g)(\widehat{\Pi}_n, u)| \\ &= \left| \mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) | \Pi_n = \widehat{\Pi}_n] - \mathbf{E}[\alpha(\widehat{\Pi}_n, \widehat{\Pi}_{n+1}, \widehat{S}_{n+1}) | \widehat{\Pi}_n] \right| \leq A + B, \end{aligned}$$

where

$$\begin{aligned} A &= \left| \mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) | \Pi_n = \widehat{\Pi}_n] - \mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) | \widehat{\Pi}_n] \right|, \\ B &= \left| \mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) - \alpha(\widehat{\Pi}_n, \widehat{\Pi}_{n+1}, \widehat{S}_{n+1}) | \widehat{\Pi}_n] \right|. \end{aligned}$$

Using the boundedness of g and v_{n+1} as well as the Lipschitz continuity of v_{n+1} given in Proposition C.10, we get a upper bound for the second term

$$\begin{aligned} B &\leq C_g \mathbf{E} \left[|\Pi_n - \widehat{\Pi}_n| | \widehat{\Pi}_n \right] + [v_{n+1}] \mathbf{E} \left[|\Pi_{n+1} - \widehat{\Pi}_{n+1}| | \widehat{\Pi}_n \right] \\ &\quad + 2C_g \mathbf{E} \left[|\mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}}| | \widehat{\Pi}_n \right]. \end{aligned} \quad (17)$$

For the first term, we use the properties of quantization as in the previous proof to obtain

$$A = \left| \mathbf{E} \left[\mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) | \Pi_n = \widehat{\Pi}_n] - \mathbf{E}[\alpha(\Pi_n, \Pi_{n+1}, S_{n+1}) | \Pi_n] \right] | \widehat{\Pi}_n \right|.$$

We now recognize operator J^m , and from Propositions C.4 and C.7, one has

$$\begin{aligned} A &= \mathbf{E} [J^m(v_{n+1}, g)(\widehat{\Pi}_n, u) - J^m(v_{n+1}, g)(\Pi_n, u) | \widehat{\Pi}_n] \\ &\leq (3C_g + 2[v_{n+1}]) \mathbf{E} \left[|\widehat{\Pi}_n - \Pi_n| | \widehat{\Pi}_n \right]. \end{aligned} \quad (18)$$

We gather the bounds provided by Eq. (17) and (18) to obtain

$$\begin{aligned} & |J(v_{n+1}, g)(\widehat{\Pi}_n, u) - \widehat{J}_{n+1}(v_{n+1}, g)(\widehat{\Pi}_n, u)| \\ &\leq (4C_g + 2[v_{n+1}]) \mathbf{E} \left[|\Pi_n - \widehat{\Pi}_n| | \widehat{\Pi}_n \right] + [v_{n+1}] \mathbf{E} \left[|\Pi_{n+1} - \widehat{\Pi}_{n+1}| | \widehat{\Pi}_n \right] \\ &\quad + 2C_g \mathbf{E} \left[|\mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}}| | \widehat{\Pi}_n \right]. \end{aligned} \quad (19)$$

Finally, combining the result for operators J and Lemma 5.13, we obtain

$$\begin{aligned} & |L^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n)| \\ &\leq [v_{n+1}] \mathbf{E} \left[|\Pi_{n+1} - \widehat{\Pi}_{n+1}| | \widehat{\Pi}_n \right] + (4C_g + 2[v_{n+1}]) \mathbf{E} \left[|\Pi_n - \widehat{\Pi}_n| | \widehat{\Pi}_n \right] \\ &\quad + 2C_g \max_{u \in Gr(\Delta)} \mathbf{E} \left[|\mathbb{1}_{\{S_{n+1} \leq u\}} - \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}}| | \widehat{\Pi}_n \right]. \end{aligned}$$

We conclude by taking the L^p norm in the equation above and using Lemma 5.12 to bound the last term

$$\begin{aligned} & \|L^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n)\|_p \\ &\leq [v_{n+1}] \|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p + (4C_g + 2[v_{n+1}]) \|\Pi_n - \widehat{\Pi}_n\|_p \\ &\quad + 2C_g (\eta^{-1} \|S_{n+1} - \widehat{S}_{n+1}\|_p + 2\eta C_\lambda), \end{aligned}$$

for some $0 < \eta < \Delta$. The best choice for η minimizing the error is when η satisfies

$$\eta^{-1} \|S_{n+1} - \widehat{S}_{n+1}\|_p = 2\eta C_\lambda,$$

which yields $\eta = (2C_\lambda)^{-1/2} (\|S_{n+1} - \widehat{S}_{n+1}\|_p)^{1/2}$. If Δ satisfies Condition (15), one has $\eta < \Delta$ as required for this optimal choice. \square

5.2.3 Fourth term of the error

Finally, the fourth error term is bounded using Lipschitz properties.

Lemma 5.15. *The fourth term Ξ_4 is bounded as follows*

$$\begin{aligned} & \|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{\Pi}_n)\|_p \\ & \leq [v_{n+1}] \|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p. \end{aligned}$$

Proof One has

$$\begin{aligned} & \|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{\Pi}_n)\|_p \\ & = \left\| \max_{m \in M} \max_{u \in Gr_m(\Delta)} \left\{ \widehat{H}_{n+1}g(\widehat{\Pi}_n, u) + \widehat{G}_{n+1}v_{n+1}(\widehat{\Pi}_n, u) \right\} \vee \widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n) \right. \\ & \quad \left. - \max_{m \in M} \max_{u \in Gr_m(\Delta)} \left\{ \widehat{H}_{n+1}g(\widehat{\Pi}_n, u) + \widehat{G}_{n+1}\widehat{v}_{n+1}(\widehat{\Pi}_n, u) \right\} \vee \widehat{K}_{n+1}\widehat{v}_{n+1}(\widehat{\Pi}_n) \right\|_p, \\ & \leq \left\| \max_{m \in M} \max_{u \in Gr_m(\Delta)} \mathbf{E} \left[\left(v_{n+1}(\widehat{\Pi}_{n+1}) - \widehat{v}_{n+1}(\widehat{\Pi}_{n+1}) \right) \mathbb{1}_{\{\widehat{S}_{n+1} \leq u\}} \middle| \widehat{\Pi}_n \right] \right. \\ & \quad \left. \vee \mathbf{E} [v_{n+1}(\widehat{\Pi}_{n+1}) - \widehat{v}_{n+1}(\widehat{\Pi}_{n+1}) | \widehat{\Pi}_n] \right\|_p \\ & \leq \|v_{n+1}(\widehat{\Pi}_{n+1}) - \widehat{v}_{n+1}(\widehat{\Pi}_{n+1})\|_p. \end{aligned} \tag{20}$$

We now introduce $v_{n+1}(\Pi_{n+1})$ to split this term into two differences. The Lipschitz continuity of v_{n+1} stated by Proposition C.10 allows us to bound the first term while we recognize V_{n+1} and \widehat{V}_{n+1} in the second one.

$$\begin{aligned} & \|\widehat{L}_{n+1}^d(v_{n+1}, g)(\widehat{\Pi}_n) - \widehat{L}_{n+1}^d(\widehat{v}_{n+1}, g)(\widehat{\Pi}_n)\|_p \\ & \leq \|v_{n+1}(\widehat{\Pi}_{n+1}) - v_{n+1}(\Pi_{n+1})\|_p + \|v_{n+1}(\Pi_{n+1}) - \widehat{v}_{n+1}(\widehat{\Pi}_{n+1})\|_p \\ & \leq [v_{n+1}] \|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p. \end{aligned}$$

Hence, the result. \square

5.3 Numerical construction of an ϵ -optimal stopping time

As in the previous section, we follow the idea of [9] and we use both the Markov chain $(\Theta_n)_{0 \leq n \leq N}$ and its quantized approximation $(\widehat{\Theta}_n)_{0 \leq n \leq N}$ to approximate the expression of the ϵ -optimal stopping time introduced in Definition 4.7. We check that we thus obtain actual stopping times for the observed filtration $(\mathfrak{F}_t^Y)_{t \geq 0}$ and that the

expected reward when stopping then is a good approximation of the value function V_0 . For all $(\pi, s) \in \mathcal{M}_1(E_0) \times \mathbb{R}^+$ and $0 \leq n \leq N$, we denote $(\hat{\pi}_n, \hat{s}_n) = \text{proj}_{\Gamma_n}(\pi, s)$. Let

$$\hat{s}_{N-n}^*(\pi, s) = \min\{t \in \text{Gr}(\Delta) : \hat{J}_n(\hat{v}_n, g)(\hat{\pi}_{n-1}, t) = \max_{u \in \text{Gr}(\Delta)} \hat{J}_n(\hat{v}_n, g)(\hat{\pi}_{n-1}, u)\}.$$

For $1 \leq n \leq N$ and $\pi \in \mathcal{M}_1(E_0)$, we define

$$\hat{r}_{N-n}(\pi, s) = \begin{cases} t_q^* & \text{if } \hat{K}_n \hat{v}_n(\hat{\pi}_{n-1}) > \max_{u \in \text{Gr}(\Delta)} \hat{J}_n(\hat{v}_n, g)(\hat{\pi}_{n-1}, u), \\ \hat{s}_{N-n}^*(\pi, s) & \text{otherwise.} \end{cases}$$

Let now for $n \geq 1$,

$$\begin{cases} \hat{R}_{n,0} &= \hat{r}_{n-1}(\Pi_0, S_0), \\ \hat{R}_{n,k} &= \hat{r}_{n-1-k}(\Pi_k, S_k) \mathbb{1}_{\{\hat{R}_{n,k-1} \geq S_k\}} \text{ for } 1 \leq k \leq n-2, \end{cases}$$

and set $\hat{U}_n = \sum_{k=1}^n \hat{R}_{n,k-1} \wedge S_k$. The following result is a direct consequence of Proposition A.11. It is a very strong result as it states that the numerically computable random variables \hat{U}_n are actual $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times.

Theorem 5.16. *For $0 \leq n \leq N$, \hat{U}_n is an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time.*

We now intend to prove that stopping at time \hat{U}_N provides a good approximation of the value function V_0 . For all $\pi \in \mathcal{M}_1(E_0)$ and $0 \leq n \leq N$ we therefore introduce the performance when abiding by the stopping rule $(\hat{U}_n)_{0 \leq n \leq N}$ and the corresponding random variables

$$\bar{v}_n(\pi) = \mathbf{E}[g(X_{\hat{U}_{N-n}}) | \Pi_0 = \pi], \quad \bar{V}_n = \bar{v}_n(\Pi_n).$$

Theorem 5.17. *Suppose that for all $0 \leq n \leq N-1$,*

$$\Delta > (2C_\lambda)^{-1/2} \|S_{n+1} - \hat{S}_{n+1}\|_p^{1/2},$$

one has then the following bound for the error between the expected reward when stopping at time \hat{U}_n and the value function

$$\begin{aligned} \|V_n - \bar{V}_n\|_p &\leq \|V_{n+1} - \bar{V}_{n+1}\|_p + \|V_n - \hat{V}_n\|_p + \|V_{n+1} - \hat{V}_{n+1}\|_p \\ &\quad + d_n \|\Pi_n - \hat{\Pi}_n\|_p + 2[v_{n+1}] \|\Pi_{n+1} - \hat{\Pi}_{n+1}\|_p \\ &\quad + b \|S_{n+1} - \hat{S}_{n+1}\|_p^{1/2}, \end{aligned}$$

where $b = 2C_g(2C_\lambda)^{1/2}$, $d_n = 7C_g + 4[v_{n+1}]$, $[v_{n+1}]$ defined in Proposition C.10.

It is important to notice that $\bar{v}_N(\pi) = \sum_{i=1}^q g(x_i) \pi^i = v_N(\pi)$ and thus $\bar{V}_N = V_N$. Therefore, the previous theorem proves that $|V_0 - \bar{V}_0|$ goes to zero when the quantization errors $(\|\Theta_n - \hat{\Theta}_n\|_p)_{0 \leq n \leq N}$ go to zero. In other words, the expected reward \bar{V}_0 when stopping at the random time \hat{U}_N can be made arbitrarily close to the value function V_0 of the partially observed optimal stopping problem (3) and hence \hat{U}_N is an ϵ -optimal stopping time.

Proof The first step consists in finding a recursion satisfied by the sequence $(\bar{V}_n)_{0 \leq n \leq N}$ in order to compare it with the dynamic programming equation giving $(\hat{V}_n)_{0 \leq n \leq N}$. Let $0 \leq n \leq N - 1$. First of all, Proposition 4.2 gives

$$\begin{aligned} & \mathbf{E}[g(X_{\hat{U}_{N-n}})|\Pi_0] \\ &= \sum_{k=0}^{N-n-1} \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_k \leq \hat{U}_{N-n}\}} \mathbb{1}_{\{\hat{R}_{N-n,k} < t_i^*\}} g \circ \Phi(x_i, \hat{R}_{N-n,k}) e^{-\Lambda(x_i, \hat{R}_{N-n,k})} \Pi_k^i | \Pi_0] \\ & \quad + \sum_{i=1}^q \mathbf{E}[\mathbb{1}_{\{T_{N-n} \leq \hat{U}_{N-n}\}} g(x_i) \Pi_n^i | \Pi_0]. \end{aligned}$$

The term corresponding to $k = 0$ in the above sum equals $Hg(\Pi_0, \hat{R}_{N-n,0})$. Taking the conditional expectation w.r.t. $\mathfrak{F}_{T_1}^Y$ in the other terms and noticing that one has $\{T_1 \leq \hat{U}_{N-n}\} = \{S_1 \leq \hat{R}_{N-n,0}\}$ yield

$$\mathbf{E}[g(X_{\hat{U}_{N-n}})|\Pi_0] = Hg(\Pi_0, \hat{R}_{N-n,0}) + \mathbf{E}[\Xi'' \mathbb{1}_{\{S_1 \leq \hat{R}_{N-n,0}\}} | \Pi_0],$$

with

$$\begin{aligned} \Xi'' &= \mathbf{E} \left[\sum_{k=1}^{N-n-1} \sum_{i=1}^q \mathbb{1}_{\{T_k \leq \hat{U}_{N-n}\}} \mathbb{1}_{\{\hat{R}_{N-n,k} < t_i^*\}} g \circ \Phi(x_i, \hat{R}_{N-n,k}) e^{-\Lambda(x_i, \hat{R}_{N-n,k})} \Pi_k^i \right. \\ & \quad \left. + \sum_{i=1}^q \mathbb{1}_{\{T_{N-n} \leq \hat{U}_{N-n}\}} g(x_i) \Pi_n^i \middle| \mathfrak{F}_{T_1}^Y \right]. \end{aligned}$$

We now make use of the Markov property of the sequence $(\Pi_n)_{n \in \mathbb{N}}$ in the term Ξ'' . Similarly to Lemma 4.8, for $n \geq 1$, on the set $\{T_1 \leq \hat{U}_{N-n}\}$, one has $\hat{R}_{N-n-1,k-1} \circ \theta = \hat{R}_{N-n,k}$ for all $1 \leq k \leq n - 1$. Thus, on the set $\{T_1 \leq \hat{U}_{N-n}\}$, one has $\hat{U}_{N-n} = T_1 + \hat{U}_{N-n-1} \circ \theta$. Recall that $\mathbb{1}_{\{T_k \leq \hat{U}_{N-n}\}} = \mathbb{1}_{\{T_{k-1} \leq \hat{U}_{N-n-1}\}} \circ \theta$. We may therefore apply the Markov property. Using Proposition 4.2, we now obtain $\Xi'' = \bar{v}_{n+1}(\Pi_1)$. Finally, we have

$$\bar{v}_n(\Pi_0) = Hg(\Pi_0, \hat{R}_{N-n,0}) + G\bar{v}_{n+1}(\Pi_0, \hat{R}_{N-n,0}) = J(\bar{v}_{n+1}, g)(\Pi_0, \hat{R}_{N-n,0}).$$

Recall that $\hat{R}_{N-n,0} = \hat{r}_{N-n-1}(\Pi_0, S_0)$ and apply the translation operator θ^n to obtain the following recursion

$$\bar{V}_n = J(\bar{v}_{n+1}, g)(\Pi_n, \hat{r}_{N-n-1}(\Pi_n, S_n)).$$

We are now able to study the error between \bar{V}_n and \hat{V}_n . Let us recall that, from its definition, $\hat{r}_{N-n-1}(\Pi_n, S_n)$ equals either $\hat{s}_{N-n-1}^*(\Pi_n, S_n)$ or t_q^* . In the latter case, notice that $J(\bar{v}_{n+1}, g)(\Pi_n, t_q^*) = K\bar{v}_{n+1}(\Pi_n)$. Eventually, one has

$$|\bar{V}_n - \hat{V}_n| \leq \mathbb{1}_{\{\hat{r}_{N-n-1}(\Pi_n, S_n) = t_q^*\}} A + \mathbb{1}_{\{\hat{r}_{N-n-1}(\Pi_n, S_n) = \hat{s}_{N-n-1}^*(\Pi_n, S_n)\}} B \quad (21)$$

with

$$\begin{cases} A &= |K\bar{v}_{n+1}(\Pi_n) - \hat{K}_{n+1}\hat{v}_{n+1}(\hat{\Pi}_n)|, \\ B &= |J(\bar{v}_{n+1}, g)(\Pi_n, \hat{s}_{N-n-1}^*(\Pi_n)) - \max_{u \in Gr(\Delta)} \hat{J}_{n+1}(\hat{v}_{n+1}, g)(\hat{\Pi}_n, u)|. \end{cases}$$

To bound the first term A , we introduce the function v_{n+1} . One has

$$\begin{aligned} A &\leq |K\bar{v}_{n+1}(\Pi_n) - Kv_{n+1}(\Pi_n)| + |Kv_{n+1}(\Pi_n) - Kv_{n+1}(\widehat{\Pi}_n)| \\ &\quad + |Kv_{n+1}(\widehat{\Pi}_n) - \widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n)| + |\widehat{K}_{n+1}v_{n+1}(\widehat{\Pi}_n) - \widehat{K}_{n+1}\widehat{v}_{n+1}(\widehat{\Pi}_n)| \\ &\leq (a) + (b) + (c) + (d). \end{aligned}$$

Let us study these four terms one by one. By definition of K , the first term (a) is bounded by $\mathbf{E} \left[|\bar{V}_{n+1} - V_{n+1}| \middle| \Pi_n \right]$. For the second term (b), we use Proposition C.8 stating the Lipschitz continuity of the operator K . The third term (c) is bounded by Lemma 5.13 and an upper bound of the fourth term (d) is given by Eq. (20). Thus, one obtains

$$\begin{aligned} A &\leq \|V_{n+1} - \bar{V}_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p + 4(C_g + [v_{n+1}])\|\Pi_n - \widehat{\Pi}_n\|_p \\ &\quad + 2[v_{n+1}]\|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p. \end{aligned}$$

We now turn to the second term B . In the following computations, denote $\widehat{s}^* = \widehat{s}_{N-n-1}^*(\Pi_n, S_n)$. Its definition yields $B = |J(\bar{v}_{n+1}, g)(\Pi_n, \widehat{s}^*) - \widehat{J}_{n+1}(\widehat{v}_{n+1}, g)(\widehat{\Pi}_n, \widehat{s}^*)|$. We split this expression into four differences again. On the set $\{\widehat{r}_{N-n-1}(\Pi_n, S_n) = \widehat{s}^*\}$, one has the equality $J(v_{n+1}, g)(\Pi_n, \widehat{s}^*) = V_{n+1}$. Hence, on this set, one obtains from Eq. (21)

$$|J(\bar{v}_{n+1}, g)(\Pi_n, \widehat{s}^*) - J(v_{n+1}, g)(\Pi_n, \widehat{s}^*)| \leq |\bar{V}_{n+1} - V_{n+1}|.$$

For the other terms, we use Propositions C.4 and C.7 for the Lipschitz continuity of J and Eq. (19) and (20) to obtain

$$\begin{aligned} B &\leq \|V_{n+1} - \bar{V}_{n+1}\|_p + \|V_{n+1} - \widehat{V}_{n+1}\|_p \\ &\quad + (7C_g + 4[v_{n+1}])\|\Pi_n - \widehat{\Pi}_n\|_p + 2[v_{n+1}]\|\Pi_{n+1} - \widehat{\Pi}_{n+1}\|_p \\ &\quad + 2C_g(2C_\lambda)^{1/2}\|S_{n+1} - \widehat{S}_{n+1}\|_p^{1/2}, \end{aligned}$$

after optimizing η . The result is obtained by taking the maximum between A and B . \square

6 Numerical example

We apply our procedure to a simple PDMP similar to the one studied in [9]. Let $E = [0; 1]$. For $x \in E$ and $t \geq 0$, the flow is defined by $\Phi(x, t) = x + vt$ so that $t^*(x) = (1 - x)/v$. We set the jump rate to $\lambda(x) = ax$ for some $a > 0$ and the transition kernel $Q(x, \cdot)$ to the uniform distribution on a finite set $E_0 \subset E$. Thus, the process evolves toward 1 and the closer it gets to 1, the more likely it will jump back to some point of E_0 . A trajectory is represented in Figure 1. The observation process is $Y_n = \varphi(Z_n) + W_n$ where $\varphi(x) = x$ and $W_n \sim \mathcal{N}(0, \sigma^2)$ for some $\sigma^2 > 0$. Finally, we choose the reward function $g(x) = x$. Our assumptions thus clearly hold. Simulations are run with $a = 3$, $v = 1$, $E_0 = \{0, 1/4, 1/2\}$, $\sigma^2 = 0.25$ and

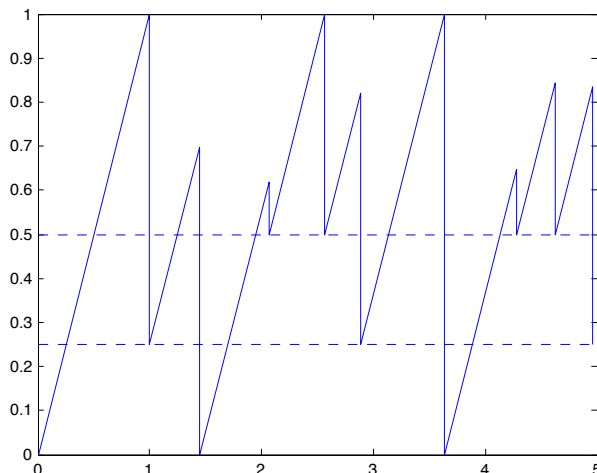


Figure 1: A trajectory of the process drawn until the 9th jump time with $a = 3$, $v = 1$ and $E_0 = \{0; \frac{1}{4}; \frac{1}{2}\}$. The dotted lines represent the possible post-jump values.

$N = 9$. The numerical approximation is implemented as follows. First, we make an exact simulator for the sequence (Z_n, S_n) . From the values of (Z_n) , one builds the observation sequence (Y_n) that allows for a recursive computation of the filter process thanks to Proposition 3.4. Thus, we can simulate trajectories of the Markov chain (Π_n, S_n) that we feed into the CLVQ algorithm to obtain quantization grids. By Monte Carlo simulations, we can also estimate the quantization errors. To run our numerical procedure, one then needs to choose the parameter Δ satisfying conditions (14) and (15). In this special case, they boil down to

$$6^{-1/2} \max_{0 \leq n \leq N-1} \|S_{n+1} - \hat{S}_{n+1}\|_p^{1/2} < \Delta < \frac{1}{8}.$$

We have chosen Δ just above the Monte Carlo approximation of the lower bound. The values are given in the second column of Table 1 for different grids sizes.

Then, we recursively compute the approximated value functions \hat{v}_n on the quantization grids. The conditional expectations are now merely weighted sums. The approximation we obtain for the value function of the partially observed optimal stopping problem are given in the fourth column of Table 1.

Finally, we implemented the construction of our ϵ -optimal stopping time and ran 10^6 Monte Carlo simulations to compute its mean performance. The results are given in the third column of Table 1.

The exact value of V_0 is unknown but one has as in [9],

$$\bar{V}_0 = \mathbf{E}[g(X_{\hat{U}_N})] \leq V_0 = \sup_{\sigma \in \Sigma_N^Y} \mathbf{E}[g(X_\sigma)] \leq \mathbf{E}\left[\sup_{0 \leq t \leq T_N} g(X_t)\right]. \quad (22)$$

Both the first and the last term may be estimated by Monte Carlo simulations. One has thus, with 10^6 trajectories, $\mathbf{E}[\sup_{0 \leq t \leq T_N} g(X_t)] = 0.9944$. The theoretical bound B_{th} of the error $|V_0 - \hat{V}_0|$ provided by Theorem 5.8 is computed using the approximated quantization errors. This bound decreases as the number of points in

Quantization grids	Δ	\bar{V}_0	\hat{V}_0	B_{em}	B_{th}
50 points	0.1179	0.7900	0.8135	0.181	683
100 points	0.0970	0.8031	0.8250	0.169	467
300 points	0.0731	0.8182	0.8407	0.154	271
500 points	0.0634	0.8250	0.8477	0.147	211
1000 points	0.0535	0.8313	0.8545	0.140	152
2000 points	0.0453	0.8361	0.8599	0.135	110
4000 points	0.0381	0.8408	0.8643	0.130	80
6000 points	0.0345	0.8430	0.8666	0.128	67
8000 points	0.0321	0.8479	0.8725	0.122	58
10000 points	0.0303	0.8497	0.8742	0.120	53
12000 points	0.0290	0.8521	0.8771	0.117	49

Table 1: Simulation results. The terms B_{em} and B_{th} respectively denote an empirical bound and the theoretical bound provided by Theorem 5.8 for the error $|V_0 - \hat{V}_0|$.

the quantization grids increases, as expected. Moreover, we computed the empirical bound given by Eq. (22) $B_{em} = \max \left\{ |\bar{V}_0 - \hat{V}_0|, |\mathbf{E}[\sup_{0 \leq t \leq T_N} g(X_t)] - \hat{V}_0| \right\}$.

A Properties of the $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times

In this section, we study the special structure of $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times.

Lemma A.1. *For all $n \in \mathbb{N}$, T_n is an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time.*

Proof Notice that for all $n \in \mathbb{N}$, $\mathbf{P}(Y_n = Y_{n+1}) = 0$. This stems from the absolute continuity of the distribution of the random variables $(W_n)_{n \in \mathbb{N}}$ since

$$\{Y_n = Y_{n+1}\} \subset \bigcup_{1 \leq i, j \leq q} \{W_n - W_{n+1} = \varphi(x_i) - \varphi(x_j)\}.$$

Hence, for all $n \in \mathbb{N}$ and $t \in \mathbb{R}^+$, one has \mathbf{P} a.s. $\{T_n \leq t\} = \{N_t \geq n\}$ where we denote $N_t = \sum_{0 \leq s \leq t} \mathbb{1}_{\{Y_s \neq Y_{s-}\}}$. The process $(N_t)_{t \geq 0}$ is \mathfrak{F}^Y -adapted thus $\{N_t \geq n\} \in \mathfrak{F}_t^Y$ and since the filtration \mathfrak{F}^Y contains the \mathbf{P} -null sets, one has $\{T_n \leq t\} \in \mathfrak{F}_t^Y$. For all $n \in \mathbb{N}$, T_n is therefore an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time. \square

We now recall Theorem A2 T33 from [6] concerning the structure of the stopping times for point processes and apply it in our case.

Definition A.2. *Define the filtration $(\mathfrak{F}_t^p)_{t \geq 0}$ as follows*

$$\mathfrak{F}_t^p = \sigma \left(\mathbb{1}_{\{Y_n \in A\}} \mathbb{1}_{\{T_n \leq s\}}; n \geq 1, 0 \leq s \leq t, A \in \mathcal{B}(\mathbb{R}^d) \right).$$

Theorem A.3. *Let σ be an $(\mathfrak{F}_t^p)_{t \geq 0}$ -stopping time. For all $n \in \mathbb{N}$, there exists a $\mathfrak{F}_{T_n}^p$ -measurable non negative random variable R_n , such that one has*

$$\sigma \wedge T_{n+1} = \left(T_n + R_n \right) \wedge T_{n+1} \quad \text{on} \quad \{\sigma \geq T_n\}.$$

Our observation process $(Y_t)_{t \geq 0}$ being a point process that fits the framework developed in [6], we apply this Theorem to $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times.

Proposition A.4. *For all $t \geq 0$, one has $\mathfrak{F}_t^Y = \mathfrak{F}_t^p$.*

Proof First prove that $\mathfrak{F}_t^Y \subset \mathfrak{F}_t^p$. Let $A \in \mathcal{B}(\mathbb{R}^d)$ and $0 \leq s \leq t$, one has

$$\{Y_s \in A\} = \bigcup_{n \in \mathbb{N}} \left(\{T_n \leq s < T_{n+1}\} \cap \{Y_n \in A\} \right) \in \mathfrak{F}_s^p \subset \mathfrak{F}_t^p.$$

Indeed, in the above equation, we used that T_0 and Y_0 are assumed to be deterministic. For the reverse inclusion, let $A \in \mathcal{B}(\mathbb{R}^d)$, $n \geq 1$ and $0 \leq s \leq t$. Recall that $Y_n = Y_{T_n}$. One has $\{Y_{T_n} \in A\} \in \mathfrak{F}_{T_n}^Y$ since $(Y_t)_{t \geq 0}$ is \mathfrak{F}^Y -adapted and T_n is an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time from Lemma A.1. Therefore, one has $\{Y_n \in A\} \cap \{T_n \leq s\} \in \mathfrak{F}_s^Y \subset \mathfrak{F}_t^Y$, showing the result. \square

We may therefore apply Theorem A.3 to $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times.

Theorem A.5. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time. For all $n \in \mathbb{N}$, there exists a non negative random variable R_n , $\mathfrak{F}_{T_n}^Y$ -measurable such that one has*

$$\sigma \wedge T_{n+1} = (T_n + R_n) \wedge T_{n+1} \quad \text{on} \quad \{\sigma \geq T_n\}.$$

We outline the following result, which is a direct consequence of the above theorem, because it will be used several times in our derivation.

Lemma A.6. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time and $(R_n)_{n \in \mathbb{N}}$ be the sequence of random variables associated to σ as introduced in Theorem A.5. For all $n \in \mathbb{N}$, $\{T_n \leq \sigma < T_{n+1}\} = \{T_n \leq \sigma\} \cap \{S_{n+1} > R_n\}$.*

Proof Theorem A.5 states that on the event $\{T_n \leq \sigma\}$, one has $\sigma \wedge T_{n+1} = T_n + (R_n \wedge S_{n+1})$ so that, still on the event $\{T_n \leq \sigma\}$, one has $(\sigma < T_{n+1}) \Leftrightarrow (R_n < S_{n+1})$. We deduce the result from this observation. \square

We now investigate the effect of the translation operator of the Markov chain $(\Pi_n, Y_n, S_n)_{n \in \mathbb{N}}$ on the $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping times. Proposition 4.1 states that $(\Pi_n, Y_n, S_n)_{n \in \mathbb{N}}$ is a $(\mathfrak{F}_{T_n}^Y)_{n \in \mathbb{N}}$ -Markov chain. Let us consider its canonical space $\Omega = (\mathcal{M}_1(E_0) \times \mathbb{R}^d \times \mathbb{R}^+)^{\mathbb{N}}$. Thus, for $\omega = (\omega_0, \omega_1, \dots) \in \Omega$, one has $(\Pi_n, Y_n, S_n)(\omega) = \omega_n$. Besides, we define the *translation operator*

$$\theta : \begin{cases} \Omega & \rightarrow \Omega \\ (\omega_0, \omega_1, \dots) & \rightarrow (\omega_1, \omega_2, \dots) \end{cases}$$

We then define $\theta^0 = Id_\Omega$ and recursively for $l \geq 2$, $\theta^l = \theta \circ \theta^{l-1}$. Thus, for all $n, l \in \mathbb{N}$, one has $(\Pi_n, Y_n, S_n) \circ \theta^l = (\Pi_{n+l}, Y_{n+l}, S_{n+l})$. As $T_0 = 0$, one has

$$T_n \circ \theta^l = \sum_{k=1}^n S_k \circ \theta^l = \sum_{k=1}^n S_{k+l} = T_{n+l} - T_l.$$

The next results of this section are given without proof because their proofs follow the very same lines as in [9] from which they are adapted. However, notice that the results from [9] cannot be applied directly to our case because the sequence $(\Pi_n, Y_n, S_n)_{n \in \mathbb{N}}$, although it is a Markov chain, is not the underlying Markov chain of some PDMP. Set now $\sigma \in \Sigma^Y$. From Theorem A.5, for all $n \in \mathbb{N}$, there exists a non negative $\mathfrak{F}_{T_n}^Y$ -measurable random variable R_n , such that, on the event $\{\sigma \geq T_n\}$, one has $\sigma \wedge T_{n+1} = (T_n + R_n) \wedge T_{n+1}$.

Lemma A.7. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time and $(R_n)_{n \in \mathbb{N}}$ be the sequence of random variables associated to σ as introduced in Theorem A.5. Let $\bar{R}_0 = R_0$ and for $k \geq 1$, $\bar{R}_k = R_k \mathbb{1}_{\{S_k \leq \bar{R}_{k-1}\}}$. One has then*

$$\sigma = \sum_{n=1}^{\infty} \bar{R}_{n-1} \wedge S_n.$$

Remark A.8. *This lemma proves that in Theorem A.5, the sequence $(R_n)_{n \in \mathbb{N}}$ can be replaced by $(\bar{R}_n)_{n \in \mathbb{N}}$. Therefore, we can assume, without loss of generality that the sequence $(R_n)_{n \in \mathbb{N}}$ satisfies the following condition: for all $n \in \mathbb{N}$, $R_{n+1} = 0$ on the event $\{S_{n+1} > R_n\}$.*

Since $\mathfrak{F}_{T_k}^Y = \sigma(Y_j, S_j, j \leq k)$ and R_k is $\mathfrak{F}_{T_k}^Y$ -measurable, there exists a sequence of real-valued measurable functions $(r_k)_{k \in \mathbb{N}}$ defined on $(\mathbb{R}^d \times \mathbb{R}^+)^{k+1}$ such that $R_k = r_k(\mathcal{G}_k)$, where $\mathcal{G}_k = (Y_0, S_0, \dots, Y_k, S_k)$.

Definition A.9. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time and $(r_n)_{n \in \mathbb{N}}$ be the sequence of functions associated to σ as introduced in Remark A.8. Let $l \geq 1$ and $(\tilde{R}_k^l)_{k \in \mathbb{N}}$ be a sequence of functions defined on $(\mathbb{R}^d \times \mathbb{R}^+)^{l+1} \times \Omega$ by $\tilde{R}_0^l(\gamma, \omega) = r_l(\gamma)$ and for $k \geq 1$, $\tilde{R}_k^l(\gamma, \omega) = r_{l+k}(\gamma, \mathcal{G}_{k-1}(\omega)) \mathbb{1}_{\{S_k \leq \tilde{R}_{k-1}^l\}}(\gamma, \omega)$.*

Proposition A.10. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time and $(\bar{R}_k)_{k \in \mathbb{N}}$ (respectively, $(\tilde{R}_k^l)_{k \in \mathbb{N}}$) be the sequence of functions associated to σ as introduced in Lemma A.7 (respectively, in Definition A.9). Assume that $T_l \leq \sigma \leq T_N$. For all $k \in \mathbb{N}$, one has then $\tilde{R}_k^l(\mathcal{G}_l, \theta^l) = \bar{R}_{l+k}$ and $\sigma = T_l + \tilde{\sigma}(\mathcal{G}_l, \theta^l)$, with $\tilde{\sigma} : (\mathbb{R}^d \times \mathbb{R}^+)^{l+1} \times \Omega \rightarrow \mathbb{R}^+$ defined as $\tilde{\sigma}(\gamma, \omega) = \sum_{n=1}^{N-l} \tilde{R}_{n-1}^l(\gamma, \omega) \wedge S_n(\omega)$.*

Proposition A.11. *Let $(U_n)_{n \in \mathbb{N}}$ be a sequence of non negative random variables such that for all n , U_n is $\mathfrak{F}_{T_n}^Y$ -measurable and $U_{n+1} = 0$ on $\{S_{n+1} > U_n\}$. We define $U = \sum_{n=1}^{\infty} U_{n-1} \wedge S_n$. Then U is an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time.*

Corollary A.12. *Let σ be an $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time and $\tilde{\sigma}$ be the mapping associated to σ introduced in Proposition A.10. For all $\gamma \in (\mathbb{R}^d \times \mathbb{R}^+)^{p+1}$, $\tilde{\sigma}(\gamma, \cdot)$ is a $(\mathfrak{F}_t^Y)_{t \geq 0}$ -stopping time.*

B Computation of a conditional expectation

The objective of this section is to prove the technical Lemma B.1 used in the proof of Proposition 4.2.

Lemma B.1. *For all $k \in \mathbb{N}$, one has $\mathbf{E}[\mathbb{1}_{\{S_{k+1} > R_k\}} | \mathfrak{F}_{T_k}^Y] = \mathbb{1}_{\{R_k < t^*(Z_k)\}} e^{-\Lambda(Z_k, R_k)}$.*

Proof First recall some results concerning the random variables $(S_k)_{k \in \mathbb{N}}$, details may be found in [8]. After a jump of the process to the point $z \in E$, the survival function of the time until the next jump is

$$\phi(t, z) = \begin{cases} 1 & \text{if } t \leq 0, \\ e^{-\Lambda(z, t)} & \text{if } 0 \leq t < t^*(z), \\ 0 & \text{if } t \geq t^*(z). \end{cases}$$

Define its generalized inverse $\psi(u, z) = \inf\{t \geq 0 \text{ such that } \phi(t, z) \leq u\}$. Then, for all $k \in \mathbb{N}$, one has $S_{k+1} = \psi(\Upsilon_k, Z_k)$, where Υ_k are i.i.d. random variables with uniform distribution on $[0; 1]$ independent from \mathfrak{F}_{T_k} . Thus, one has $\mathbf{E}[\mathbb{1}_{\{S_{k+1} > R_k\}} | \mathfrak{F}_{T_k}] = \mathbf{E}[f(\Upsilon_k, Z_k, R_k) | \mathfrak{F}_{T_k}]$ where $f(u, z, r) = \mathbb{1}_{\{\psi(u, z) > r\}}$. As (Z_k, R_k) is \mathfrak{F}_{T_k} -measurable, Υ_k is independent from \mathfrak{F}_{T_k} and $\mathbf{E}[\mathbb{1}_{\{\psi(\Upsilon_k, z) > r\}}] = \mathbb{1}_{\{r < t^*(z)\}} e^{-\Lambda(z, r)}$, [15, Proposition 11.2] yields the result. \square

C Lipschitz properties

In this section, we derive the Lipschitz properties of our operators in order to obtain them for the value functions $(v_n)_{0 \leq n \leq N}$. Similarly to the proof of Proposition 4.1, we first derive the integral form of operators G and H .

Lemma C.1. *For all $h \in B(E)$, $v \in B(\mathcal{M}_1(E_0))$ and $(\pi, u) \in \mathcal{M}_1(E_0) \times \mathbb{R}^+$, one has*

$$\begin{aligned} Gv(\pi, u) &= Iv(\pi, u) + \sum_{i=1}^q \pi^i \mathbb{1}_{\{t_i^* \leq u\}} e^{-\Lambda(x_i, t_i^*)} \\ &\quad \times \int_{\mathbb{R}^d} v(\Psi(\pi, y', t_i^*)) \sum_{j=1}^q Q(\Phi(x_i, t_i^*), x_j) f_W(y' - \varphi(x_j)) dy', \\ Hh(\pi, u) &= \sum_{i=1}^q \pi^i \mathbb{1}_{\{u < t_i^*\}} e^{-\Lambda(x_i, u)} h \circ \Phi(x_i, u), \end{aligned}$$

where

$$\begin{aligned} Iv(\pi, u) &= \sum_{i=1}^q \pi^i \int_0^{u \wedge t_i^*} \left(\lambda \circ \Phi(x_i, s') e^{-\Lambda(x_i, s')} \right. \\ &\quad \left. \times \int_{\mathbb{R}^d} v(\Psi(\pi, y', s')) \sum_{j=1}^q Q(\Phi(x_i, s'), x_j) f_W(y' - \varphi(x_j)) dy' \right) ds'. \end{aligned}$$

Now, notice that the functions $Hh(\pi, \cdot)$ and $Gv(\pi, \cdot)$ are not continuous. However, they are càdlàg with a finite number of jumps. Therefore, they can be rewritten as sums of continuous functions as follows.

Definition C.2. *For all $m \in \{0, \dots, q-1\}$, we define the operators $G^m: B(\mathcal{M}_1(E_0)) \rightarrow B(\mathcal{M}_1(E_0) \times \mathbb{R}^+)$ and $H^m: B(E) \rightarrow B(\mathcal{M}_1(E_0) \times \mathbb{R}^+)$ as follows*

- if $u < t_m^*$, $G^m v(\pi, u) = Gv(\pi, t_m^*)$ and $H^m h(\pi, u) = Hh(\pi, t_m^*)$,

- if $u \geq t_m^*$,

$$\begin{aligned}
G^m v(\pi, u) &= Iv(\pi, u \wedge t_{m+1}^*) + \sum_{i=1}^m \pi^i e^{-\Lambda(x_i, t_i^*)} \\
&\quad \times \int_{\mathbb{R}^d} v(\Psi(\pi, y', t_i^*)) \sum_{j=1}^q Q(\Phi(x_i, t_i^*), x_j) f_W(y' - \varphi(x_j)) dy', \\
H^m h(\pi, u) &= \sum_{i=m+1}^q \pi^i e^{-\Lambda(x_i, u \wedge t_{m+1}^*)} h \circ \Phi(x_i, u \wedge t_{m+1}^*).
\end{aligned}$$

We also define $J^m(v, h)(\pi, u) = H^m h(\pi, u) + G^m v(\pi, u)$.

Remark C.3. For all $m \in \{0, \dots, q-1\}$ and for all $h \in B(E)$, $v \in B(\mathcal{M}_1(E_0))$ and $(\pi, u) \in \mathcal{M}_1(E_0) \times \mathbb{R}^+$, the functions $u \rightarrow G^m v(\pi, u)$, $u \rightarrow H^m h(\pi, u)$ and $u \rightarrow J^m(v, h)(\pi, u)$ are continuous. Moreover, they are constant on $[0; t_m^*]$ and on $[t_{m+1}^*; +\infty[$ and one has

$$\begin{aligned}
Gv(\pi, u) &= \sum_{m=0}^{q-1} \mathbb{1}_{[t_m^*, t_{m+1}^*]}(u) G^m v(\pi, u), \\
Hh(\pi, u) &= \sum_{m=0}^{q-1} \mathbb{1}_{[t_m^*, t_{m+1}^*]}(u) H^m h(\pi, u), \\
J(v, h)(\pi, u) &= \sum_{m=0}^{q-1} \mathbb{1}_{[t_m^*, t_{m+1}^*]}(u) J^m(v, h)(\pi, u).
\end{aligned}$$

We now investigate the Lipschitz properties of our operators.

Proposition C.4. For $m \in M$, $((\pi, u), (\tilde{\pi}, \tilde{u})) \in (\mathcal{M}_1(E_0) \times \mathbb{R}^+)^2$, one has

$$|H^m g(\pi, u) - H^m g(\tilde{\pi}, \tilde{u})| \leq C_g |\pi - \tilde{\pi}| + ([g]_2 + C_g C_\lambda) |u - \tilde{u}|.$$

Proof Since the function $u \rightarrow H^m h(\pi, u)$ is constant on the intervals $[0; t_m^*]$ and $[t_{m+1}^*; +\infty[$, we may assume that $u, \tilde{u} \in [t_m^*; t_{m+1}^*]$ so that one has $H^m g(\pi, u) = \sum_{i=m+1}^q \pi^i e^{-\Lambda(x_i, u)} g \circ \Phi(x_i, u)$, and similarly for $H^m g(\tilde{\pi}, \tilde{u})$. Then, on the one hand, one has

$$\begin{aligned}
|H^m g(\pi, u) - H^m g(\tilde{\pi}, \tilde{u})| &= \left| \sum_{i=m+1}^q (\pi^i - \tilde{\pi}^i) e^{-\Lambda(x_i, u)} g \circ \Phi(x_i, u) \right| \\
&\leq C_g \sum_{i=m+1}^q |\pi^i - \tilde{\pi}^i|.
\end{aligned}$$

On the other hand, Lemma A.1 in [9] yields

$$|e^{-\Lambda(x_i, u)} g \circ \Phi(x_i, u) - e^{-\Lambda(x_i, \tilde{u})} g \circ \Phi(x_i, \tilde{u})| \leq ([g]_2 + C_g C_\lambda) |u - \tilde{u}|,$$

showing the result. \square

The following technical lemma will be useful to derive the Lipschitz properties of the operator I . The first part of its proof is adapted from [18].

Lemma C.5. For all $\pi, \tilde{\pi} \in \mathcal{M}_1(E_0)$ and $m \in M$, one has

$$\sum_{m=0}^{q-1} \int_{t_m^*}^{t_{m+1}^*} \int_{\mathbb{R}^d} |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')| \bar{\Psi}_m(\pi, y', s') dy' ds' \leq 2|\pi - \tilde{\pi}|.$$

Proof Let $s' \in]t_m^*; t_{m+1}^*]$ and $y' \in \mathbb{R}^d$. In the following computation, we denote $\tau = (\pi, y', s')$ and $\tilde{\tau} = (\tilde{\pi}, y', s')$, one has

$$\begin{aligned} |\Psi(\tau) - \Psi(\tilde{\tau})| \bar{\Psi}_m(\tau) &= \sum_{j=1}^q \left| \frac{\Psi_m^j(\tau)}{\bar{\Psi}_m(\tau)} - \frac{\Psi_m^j(\tilde{\tau})}{\bar{\Psi}_m(\tilde{\tau})} \right| \bar{\Psi}_m(\tau) \\ &= \sum_{j=1}^q \left| \frac{\Psi_m^j(\tau) \bar{\Psi}_m(\tilde{\tau}) - \Psi_m^j(\tilde{\tau}) \bar{\Psi}_m(\tau)}{\bar{\Psi}_m(\tilde{\tau})} \right| \\ &\leq \sum_{j=1}^q |\Psi_m^j(\tau) - \Psi_m^j(\tilde{\tau})| + \sum_{j=1}^q \frac{\Psi_m^j(\tilde{\tau})}{\bar{\Psi}_m(\tilde{\tau})} |\bar{\Psi}_m(\tau) - \bar{\Psi}_m(\tilde{\tau})|. \end{aligned}$$

Notice that $\sum_{j=1}^q \Psi_m^j(\tilde{\tau}) = \bar{\Psi}_m(\tilde{\tau})$ so that the second sum above reduces to $|\bar{\Psi}_m(\tau) - \bar{\Psi}_m(\tilde{\tau})| = \sum_{j=1}^q |\Psi_m^j(\tau) - \Psi_m^j(\tilde{\tau})|$. Finally, one has

$$|\Psi(\tau) - \Psi(\tilde{\tau})| \bar{\Psi}_m(\tau) \leq 2 \sum_{j=1}^q |\Psi_m^j(\tau) - \Psi_m^j(\tilde{\tau})|.$$

As $\int_{\mathbb{R}^d} f_W(y' - \varphi(x_j)) dy' = 1$ and $\sum_{j=1}^q Q(\Phi(x_i, s'), x_j) = 1$, one obtains

$$\begin{aligned} &\sum_{m=0}^{q-1} \int_{t_m^*}^{t_{m+1}^*} \int_{\mathbb{R}^d} |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')| \bar{\Psi}_m(\pi, y', s') dy' ds' \\ &\leq 2 \sum_{m=0}^{q-1} \int_{t_m^*}^{t_{m+1}^*} \sum_{j=1}^q \int_{\mathbb{R}^d} |\Psi_m^j(\pi, y', s') - \Psi_m^j(\tilde{\pi}, y', s')| dy' ds' \\ &\leq 2 \sum_{m=0}^{q-1} \sum_{i=m+1}^q \int_{t_m^*}^{t_{m+1}^*} \sum_{j=1}^q \int_{\mathbb{R}^d} |\pi^i - \tilde{\pi}^i| \lambda(\Phi(x_i, s')) e^{-\Lambda(x_i, s')} \\ &\quad \times Q(\Phi(x_i, s'), x_j) f_W(y' - \varphi(x_j)) dy' ds' \\ &\leq 2 \sum_{m=0}^{q-1} \sum_{i=m+1}^q |\pi^i - \tilde{\pi}^i| \int_{t_m^*}^{t_{m+1}^*} \lambda(\Phi(x_i, s')) e^{-\Lambda(x_i, s')} ds' \\ &\leq 2 \sum_{i=1}^q |\pi^i - \tilde{\pi}^i| \int_0^{t_i^*} \lambda(\Phi(x_i, s')) e^{-\Lambda(x_i, s')} ds'. \end{aligned}$$

We obtain the result as $\int_0^{t_i^*} \lambda(\Phi(x_i, s')) e^{-\Lambda(x_i, s')} ds' = 1 - e^{-\Lambda(x_i, t_i^*)} \leq 1$. \square

Proposition C.6. For $v \in BL(\mathcal{M}_1(E_0))$ and $((\pi, u), (\tilde{\pi}, \tilde{u})) \in (\mathcal{M}_1(E_0) \times \mathbb{R}^+)^2$, one has

$$|Iv(\pi, u) - Iv(\tilde{\pi}, \tilde{u})| \leq (C_v + 2[v])|\pi - \tilde{\pi}| + C_v C_\lambda |u - \tilde{u}|.$$

Proof On the one hand, one clearly has

$$|Iv(\pi, u) - Iv(\pi, \tilde{u})| \leq \sum_{i=1}^q \pi^i |u \wedge t_i^* - \tilde{u} \wedge t_i^*| C_v C_\lambda \leq C_v C_\lambda |u - \tilde{u}|.$$

On the other hand, one has

$$\begin{aligned} & |Iv(\pi, u) - Iv(\tilde{\pi}, u)| \\ & \leq C_v |\pi - \tilde{\pi}| + \sum_{i=1}^q \pi^i \int_0^{t_i^*} \int_{\mathbb{R}^d} |v(\Psi(\pi, y', s')) - v(\Psi(\tilde{\pi}, y', s'))| \\ & \quad \times \sum_{j=1}^q Q(\Phi(x_i, s'), x_j) f_W(y' - \varphi(x_j)) \lambda \circ \Phi(x_i, s') e^{-\Lambda(x_i, s')} dy' ds'. \end{aligned}$$

Besides, we have assumed that v is Lipschitz continuous so that one has

$$|v(\Psi(\pi, y', s')) - v(\Psi(\tilde{\pi}, y', s'))| \leq [v] |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')|.$$

Thus, one has

$$\begin{aligned} & |Iv(\pi, y, s, u) - Iv(\tilde{\pi}, y, s, u)| \\ & \leq C_v |\pi - \tilde{\pi}| + [v] \sum_{i=1}^q \pi^i \int_0^{t_i^*} \int_{\mathbb{R}^d} |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')| \\ & \quad \sum_{j=1}^q Q(\Phi(x_i, s'), x_j) f_W(y' - \varphi(x_j)) \lambda \circ \Phi(x_i, s') e^{-\Lambda(x_i, s')} dy' ds' \\ & \leq C_v |\pi - \tilde{\pi}| + [v] \sum_{m=0}^{q-1} \sum_{i=m+1}^q \pi^i \int_{t_m^*}^{t_{m+1}^*} \int_{\mathbb{R}^d} |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')| \\ & \quad \times \sum_{j=1}^q Q(\Phi(x_i, s'), x_j) f_W(y' - \varphi(x_j)) \lambda \circ \Phi(x_i, s') e^{-\Lambda(x_i, s')} dy' ds' \\ & \leq C_v |\pi - \tilde{\pi}| + [v] \sum_{m=0}^{q-1} \int_{t_m^*}^{t_{m+1}^*} \int_{\mathbb{R}^d} |\Psi(\pi, y', s') - \Psi(\tilde{\pi}, y', s')| \bar{\Psi}_m(\pi, y', s') dy' ds'. \end{aligned}$$

The previous lemma provides the result. \square

Proposition C.7. For $m \in M$, $v \in BL(\mathcal{M}_1(E_0))$ and $((\pi, u), (\tilde{\pi}, \tilde{u})) \in (\mathcal{M}_1(E_0) \times \mathbb{R}^+)^2$, one has

$$|G^m v(\pi, u) - G^m v(\tilde{\pi}, \tilde{u})| \leq (2C_v + 2[v]) |\pi - \tilde{\pi}| + C_v C_\lambda |u - \tilde{u}|.$$

Proof As in the proof of Proposition C.4, we may assume without loss of generality that $u, \tilde{u} \in [t_m^*; t_{m+1}^*]$ so that one has

$$\begin{aligned} G^m v(\pi, u) &= Iv(\pi, u) + \sum_{i=1}^m \pi^i e^{-\Lambda(x_i, t_i^*)} \\ & \quad \times \sum_{j=1}^q \int_{\mathbb{R}^d} v(\Psi(\pi, y', t_i^*)) Q(\Phi(x_i, t_i^*), x_j) f_W(y' - \varphi(x_j)) dy', \end{aligned}$$

and similarly for $G^m v(\tilde{\pi}, \tilde{u})$. The second term does not depend on u thus

$$\begin{aligned} |G^m v(\pi, u) - G^m v(\pi, \tilde{u})| &= |Iv(\pi, u) - Iv(\pi, \tilde{u})| \\ &\leq |Iv(\pi, u) - Iv(\tilde{\pi}, u)| + C_v |\pi - \tilde{\pi}|, \end{aligned}$$

as $\Psi(\pi, y', t_i^*) = \Psi(\tilde{\pi}, y', t_i^*)$ by Proposition 3.4. This yields the result. \square

Proposition C.8. *For all $v \in BL(\mathcal{M}_1(E_0))$ and $(\pi, \tilde{\pi}) \in \mathcal{M}_1(E_0)^2$, one has*

$$|Kv(\pi) - Kv(\tilde{\pi})| \leq (2C_v + 2[v])|\pi - \tilde{\pi}|.$$

Proof As $Kv(\pi) = Gv(\pi, t_q^*)$, this is a consequence of Proposition C.7. \square

Proposition C.9. *For $v \in BL(\mathcal{M}_1(E_0))$ and $(\pi, \tilde{\pi}) \in \mathcal{M}_1(E_0)^2$, one has*

$$|L(v, g)(\pi) - L(v, g)(\tilde{\pi})| \leq (C_g + 2C_v + 2[v])|\pi - \tilde{\pi}|.$$

Proof One has

$$\begin{aligned} &|L(v, g)(\pi) - L(v, g)(\tilde{\pi})| \\ &\leq \max_{m \in M} \left\{ \sup_{u \in [t_m^*, t_{m+1}^*]} |J^m(v, g)(\pi, u) - J^m(v, g)(\tilde{\pi}, u)| \right\} \vee |Kv(\pi) - Kv(\tilde{\pi})| \\ &\leq (C_g + 2C_v + 2[v])|\pi - \tilde{\pi}|, \end{aligned}$$

using Propositions C.4, C.7 and C.8 since $J^m(v, g) = H^m g + G^m v$. \square

Proposition C.10. *For all $n \in \{0, \dots, N\}$, one has $v_n \in BL(\mathcal{M}_1(E_0))$ with $C_{v_n} \leq C_g$ and $[v_n] \leq (2^{N-n+2} - 3)C_g$.*

Proof We proved that v_n is the value function of the optimal stopping problem with horizon T_{N-n} thus one has $v_n(\pi) = \sup_{\sigma \in \Sigma_{N-n}^Y} \mathbf{E}[g(X_\sigma) | \Pi_0 = \pi] \leq C_g$. Therefore v_n is bounded and $C_{v_n} \leq C_g$. The second assessment is proved by backward induction. Let $\pi, \tilde{\pi} \in \mathcal{M}_1(E_0)$. One has

$$|v_N(\pi) - v_N(\tilde{\pi})| \leq \sum_{j=1}^N g(x_j) |\pi^j - \tilde{\pi}^j| \leq C_g |\pi - \tilde{\pi}|.$$

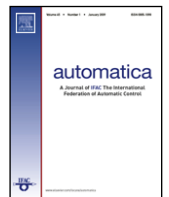
Therefore, we have the result for $n = N$ with $[v_N] \leq C_g$. Moreover, since $v_n = L(v_{n+1}, g)$ for $0 \leq n \leq N - 1$, Proposition C.9 yields $[v_n] \leq 3C_g + 2[v_{n+1}]$ which proves the propagation of the induction. \square

References

- [1] E. Arjas, P. Haara, and I. Norros. Filtering the histories of a partially observed marked point process. *Stochastic Process. Appl.*, 40(2):225–250, 1992.
- [2] V. Bally and G. Pagès. A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6):1003–1049, 2003.
- [3] V. Bally, G. Pagès, and J. Printems. A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance*, 15(1):119–168, 2005.
- [4] N. Bäuerle and U. Rieder. *Markov decision processes with applications to finance*. Universitext. Springer, Heidelberg, 2011.
- [5] A. Brandejsky, B. de Saporta, and F. Dufour. Numerical method for expectations of piecewise-deterministic markov processes. *CAMCoS*, 7(1):63–104, 2012.
- [6] P. Brémaud. *Point processes and queues*. Springer Series in Statistics. Springer-Verlag, New York, 1981.
- [7] D. Chafaï, F. Malrieu, and K. Paroux. On the long time behavior of the TCP window size process. *Stochastic Process. Appl.*, 120(8):1518–1534, 2010.
- [8] M. H. A. Davis. *Markov models and optimization*, volume 49 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1993.
- [9] B. de Saporta, F. Dufour, and K. Gonzalez. Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Probab.*, 20(5):1607–1637, 2010.
- [10] B. de Saporta, F. Dufour, H. Zhang, and C. Elegbede. Optimal stopping for the predictive maintenance of a structure subject to corrosion. *Journal of Risk and Reliability*, 226(2):169–181, 2012.
- [11] E. Edoli and W. J. Runggaldier. On optimal investment in a reinsurance context with a point process market model. *Insurance Math. Econom.*, 47(3):315–326, 2010.
- [12] R. M. Gray and D. L. Neuhoff. Quantization. *IEEE Trans. Inform. Theory*, 44(6):2325–2383, 1998. Information theory: 1948–1998.
- [13] U. S. Gugerli. Optimal stopping of a piecewise-deterministic Markov process. *Stochastics*, 19(4):221–236, 1986.
- [14] M. Ludkovski and S. O. Sezer. Finite horizon decision timing with partially observable Poisson processes. *Stoch. Models*, 28(2):207–247, 2012.
- [15] J.-Y. Ouvrard. *Probabilités 2*. Enseignement des mathématiques. Cassini, 2004.

- [16] G. Pagès, H. Pham, and J. Printems. Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance*, pages 253–297. Birkhäuser Boston, Boston, MA, 2004.
- [17] K. Pakdaman, M. Thiellen, and G. Wainrib. Fluid limit theorems for stochastic hybrid systems with application to neuron models. *Adv. in Appl. Probab.*, 42(3):761–794, 2010.
- [18] H. Pham, W. Runggaldier, and A. Sellami. Approximation by quantization of the filter process and applications to optimal stopping problems under partial observation. *Monte Carlo Methods Appl.*, 11(1):57–81, 2005.

- 2.6 [dSD12] B. DE SAPORTA AND F. DUFOUR. Numerical method for impulse control of piecewise deterministic markov processes. *Automatica*, 48 :779–793, 2012.



Numerical method for impulse control of piecewise deterministic Markov processes[☆]

Benoîte de Saporta^{a,b,c}, François Dufour^{b,c}

^a Université de Bordeaux, GREThA CNRS UMR 5113, Avenue Léon Duguit, 33608 Pessac Cedex, France

^b Université de Bordeaux, IMB CNRS UMR 5251, 351 cours de la Libération, 33405 Talence Cedex, France

^c INRIA Bordeaux Sud Ouest team CQFD, 351 cours de la Libération, 33405 Talence Cedex, France

ARTICLE INFO

Article history:

Received 30 November 2010

Received in revised form

26 May 2011

Accepted 7 October 2011

Available online 19 March 2012

Keywords:

Impulse control

Piecewise deterministic Markov processes

Long-run average cost

Numerical approximation

Quantization

ABSTRACT

This paper presents a numerical method to calculate the value function for a general discounted impulse control problem for piecewise deterministic Markov processes. Our approach is based on a quantization technique for the underlying Markov chain defined by the post jump location and inter-arrival time. Convergence results are obtained and more importantly we are able to give a convergence rate of the algorithm. The paper is illustrated by a numerical example.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

We present here a numerical method to compute the value function of an impulse control problem for a piecewise deterministic Markov process. Our approach is based on the quantization of an underlying discrete-time Markov chain related to the continuous-time process and path-adapted time discretization grids.

Piecewise-deterministic Markov processes (PDMPs) have been introduced in the literature by Davis (1993) as a general class of stochastic hybrid models. PDMPs are a family of Markov processes involving deterministic motion punctuated by random jumps. The motion of the PDMP includes both continuous and discrete variables $\{(X(t), \mathcal{Y}(t))\}$. The hybrid state space (continuous/discrete) is defined as $\mathbb{R}^d \times M$ where M is a countable set. The process depends on three local characteristics, namely the flow ϕ , the jump rate λ and the transition measure Q , which specifies the post-jump location. Starting from $(x, \nu) \in \mathbb{R}^d \times M$ the motion of the process follows the trajectory $(\phi_\nu(x, t), \nu)$

until the first jump time T_1 which occurs either spontaneously in a Poisson-like fashion with rate $\lambda_\nu(\phi_\nu(x, t))$ or when the flow $\phi_\nu(x, t)$ hits the boundary of the state-space. In either case the location of the process at the jump time $T_1: (X(T_1), \mathcal{Y}(T_1)) = (Z_1, y_1)$ is selected by the transition measure $Q_\nu(\phi_\nu(x, T_1), \cdot)$. Starting from (Z_1, y_1) , we now select the next inter-jump time $T_2 - T_1$ and post jump location $(X(T_2), \mathcal{Y}(T_2)) = (Z_2, y_2)$. This gives a piecewise deterministic trajectory for $\{(X(t), \mathcal{Y}(t))\}$ with jump times $\{T_k\}$ and post jump locations $\{(Z_k, y_k)\}$ which follows the flow ϕ between two jumps. A suitable choice of the state space and the local characteristics ϕ , λ , and Q provides stochastic models covering a great number of problems of operations research; see Davis (1993). To simplify notation, there is no loss of generality in considering that the state space of the PDMP is taken simply as a subset of \mathbb{R}^d rather than a product space $\mathbb{R}^d \times M$ as described above; see Remark 24.9 in Davis (1993) for details.

An impulse control strategy consists of a sequence of single interventions introducing a jump of the process at some controller-specified stopping time and moving the process at that time to some new point in the state space. Our impulse control problem consists in choosing a strategy (if it exists) that minimizes the expected sum of discounted running and intervention costs up to infinity, and computing the optimal cost thus achieved. Many applied problems fall into this class, such as inventory problems in which a sequence of restocking decisions is made, or optimal maintenance of complex systems with components subject to failure and repair.

[☆] This work was supported by ARPEGE programme of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004. The material in this paper was presented at the 18th IFAC world congress, August 28–September 2 2011, Milano, Italy. This paper was recommended for publication in revised form by Associate Editor George Yin under the direction of Editor Ian R. Petersen.

E-mail addresses: saporta@math.u-bordeaux1.fr (B. de Saporta), dufour@math.u-bordeaux1.fr (F. Dufour).

Impulse control problems of PDMPs in the context of an expected discounted cost have been considered in Costa and Davis (1989), Dempster and Ye (1995), Gaćtarek (1991, 1992), Lenhart (1989). Roughly speaking, in Costa and Davis (1989) the authors study this impulse control problem by using the value improvement approach while in Dempster and Ye (1995), Gaćtarek (1991, 1992), Lenhart (1989) the authors choose to analyse it by using the variational inequality approach. In Costa and Davis (1989), the authors also consider a numerical procedure. By showing that iteration of the single-jump-or-intervention operator generates a sequence of functions converging to the value function of the problem, they derive an algorithm to compute an approximation of that value function. Their approach is also based on a uniform discretization of the state space similar to the one proposed by Kushner (1977). In particular, they derive a convergence result for the approximation scheme but no estimation of the rate of convergence is given. To the best of our knowledge, it is the only paper presenting a computational method for solving the impulse control problem for a PDMP in the context of discounted cost. Remark that a similar procedure has been applied by Costa (1993) to derive a numerical scheme for the impulse control problem with a long run average cost.

Our approach is also based on the iteration of the single-jump-or-intervention operator, but we want to derive a convergence rate for our approximation. Our method does not rely on a *blind* discretization of the state space, but on a discretization that depends on time and takes into account the random nature of the process. Our approach involves a quantization procedure. Roughly speaking, quantization is a technique that approximates a continuous state space random variable X by a random variable \hat{X} taking only finitely many values and such that the difference between X and \hat{X} is minimal for the L_p norm. Quantization methods have been developed recently in numerical probability, nonlinear filtering or optimal stochastic control with applications in finance, see e.g. Bally and Pagès (2003), Bally, Pagès, and Printems (2005), Pagès (1998), Pagès and Pham (2005), Pagès, Pham, and Printems (2004a,b) and references therein. It has also been successfully used by the authors to compute an approximation of the value function and optimal strategy for the optimal stopping problem for PDMPs in de Saporta, Dufour, and Gonzalez (2010).

Although the value function of the impulse control problem can be computed by iterating implicit optimal stopping problems, see Costa and Davis (1989) Proposition 2 or Davis (1993) Proposition 54.18, from a numerical point of view the impulse control is much more difficult to handle than the optimal stopping problem. Indeed, for the optimal stopping problem, the value function is computed as the limit of a sequence (v_n) constructed by iterating an operator L . This iteration procedure yields an iterative construction of a sequence of random variables $v_n(Z_n)$ (where (Z_n) is an embedded discrete-time process). This was the keystone of our approximation procedure. As regards impulse control, the iterative construction for the corresponding random variables does not hold anymore, see Section 4 for details. This is mostly due to the fact that not only does the controller choose times to stop the process, but they also choose a new starting point for the process to restart from after each intervention. This makes the single-jump-or-intervention operator significantly more complicated to iterate than the single-jump-or-stop operator used for optimal stopping. We manage to overcome this extra difficulty by using two series of quantization grids instead of just the one we used for optimal stopping.

The paper is organized as follows. In Section 2 we give a precise definition of a PDMP and state our notation and assumptions. In Section 4, we present the impulse control problem and recall the iterative construction of the value function presented in Costa and Davis (1989). In Section 5, we explain our approximation procedure and prove its convergence with error bounds. Finally in Section 6 we present a numerical example. Some technical results are postponed to the Appendix.

2. Definitions and assumptions

We first give a precise definition of a piecewise deterministic Markov process (PDMP). Some general assumptions are presented in the end of this section. Let us introduce first some standard notation. Let M be a metric space. $\mathbf{B}(M)$ is the set of real-valued, bounded, measurable functions defined on M . The Borel σ -field of M is denoted by $\mathcal{B}(M)$. Let Q be a Markov kernel on $(M, \mathcal{B}(M))$ and $w \in \mathbf{B}(M)$, $Qw(x) = \int_M w(y)Q(x, dy)$ for $x \in M$. For $(a, b) \in \mathbb{R}^2$, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$. Let E be an open subset of \mathbb{R}^d , ∂E its boundary and \bar{E} its closure. A PDMP is determined by its local characteristics (ϕ, λ, Q) where:

- the flow $\phi: \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is a one-parameter group of homeomorphisms: ϕ is continuous, $\phi(\cdot, t)$ is an homeomorphism for each $t \in \mathbb{R}$ satisfying $\phi(\cdot, t+s) = \phi(\phi(\cdot, s), t)$. For all x in E , let us denote

$$t^*(x) \doteq \inf\{t > 0: \phi(x, t) \in \partial E\},$$

with the convention $\inf \emptyset = \infty$.

- the jump rate $\lambda: \bar{E} \rightarrow \mathbb{R}_+$ is assumed to be a measurable function.
- Q is a Markov kernel on $(\bar{E}, \mathcal{B}(\bar{E}))$ satisfying the following property:

$$(\forall x \in \bar{E}), \quad Q(x, E - \{x\}) = 1.$$

From these characteristics, it can be shown, see Davis (1993, pp. 62–66), that there exists a filtered probability space $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \{\mathbf{P}_x\}_{x \in E})$ such that the motion of the process $\{X(t)\}$ starting from a point $x \in E$ may be constructed as follows. Take a random variable T_1 such that

$$\mathbf{P}_x(T_1 > t) \doteq \begin{cases} e^{-\Lambda(x,t)} & \text{for } t < t^*(x), \\ 0 & \text{for } t \geq t^*(x), \end{cases}$$

where for $x \in E$ and $t \in [0, t^*(x)]$

$$\Lambda(x, t) \doteq \int_0^t \lambda(\phi(x, s)) ds.$$

If T_1 generated according to the above probability is equal to infinity, then for $t \in \mathbb{R}_+$, $X(t) = \phi(x, t)$. Otherwise select independently an E -valued random variable (labelled Z_1) having distribution $Q(\phi(x, T_1), \cdot)$, namely $\mathbf{P}_x(Z_1 \in A) = Q(\phi(x, T_1), A)$ for any $A \in \mathcal{B}(E)$. The trajectory of $\{X(t)\}$ starting at x , for $t \leq T_1$, is given by

$$X(t) \doteq \begin{cases} \phi(x, t) & \text{for } t < T_1, \\ Z_1 & \text{for } t = T_1. \end{cases}$$

Starting from $X(T_1) = Z_1$, we now select the next inter-jump time $T_2 - T_1$ and post-jump location $X(T_2) = Z_2$ is a similar way.

This gives a strong Markov process $\{X(t)\}$ with jump times $\{T_k\}_{k \in \mathbb{N}}$ (where $T_0 = 0$). Associated to $\{X(t)\}$, there exists a discrete time process $(\Theta_n)_{n \in \mathbb{N}}$ defined by $\Theta_n = (Z_n, S_n)$ with $Z_n = X(T_n)$ and $S_n = T_n - T_{n-1}$ for $n \geq 1$ and $S_0 = 0$. Clearly, the process $(\Theta_n)_{n \in \mathbb{N}}$ is a Markov chain, and it is the only source of randomness of the process.

We define the following space of functions continuous along the flow with limit towards the boundary:

$$\mathbf{C} = \left\{ w \in \mathbf{B}(E): w(\phi(x, \cdot)): [0, t^*(x)] \mapsto \mathbb{R} \text{ is continuous for each } x \in E \text{ and whenever } t^*(x) < \infty \text{ the limit } \lim_{t \rightarrow t^*(x)} w(\phi(x, t)) \text{ exists} \right\}.$$

For $w \in \mathbf{C}$, we define $w(\phi(x, t^*(x)))$ by the limit $\lim_{t \rightarrow t^*(x)} w(\phi(x, t))$ (note that the limit exists by assumption). Let us introduce \mathbf{L} as the set of functions $w \in \mathbf{C}$ satisfying the following properties:

1. there exists $[w]_1 \in \mathbb{R}_+$ such that for any $(x, y) \in E^2, u \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|w(\phi(x, u)) - w(\phi(y, u))| \leq [w]_1 |x - y|,$$
2. there exists $[w]_2 \in \mathbb{R}_+$ such that for any $x \in E$, and $(t, s) \in [0, t^*(x)]^2$, one has

$$|w(\phi(x, t)) - w(\phi(x, s))| \leq [w]_2 |t - s|,$$
3. there exists $[w]_* \in \mathbb{R}_+$ such that for any $(x, y) \in E^2$, one has

$$|w(\phi(x, t^*(x))) - w(\phi(y, t^*(y)))| \leq [w]_* |x - y|.$$

In the sequel, for any function w in \mathbf{C} , we denote by C_w its bound: $C_w = \sup_{x \in E} |w(x)|$.

The following assumptions will be in force throughout.

Assumption 2.1. The jump rate λ is bounded and there exists $[\lambda]_1 \in \mathbb{R}_+$ such that for any $(x, y) \in E^2, u \in [0, t^*(x) \wedge t^*(y)]$,

$$|\lambda(\phi(x, u)) - \lambda(\phi(y, u))| \leq [\lambda]_1 |x - y|.$$

Assumption 2.2. The exit time t^* is bounded and Lipschitz-continuous on E .

Assumption 2.3. The Markov kernel Q is Lipschitz in the following sense: there exists $[Q] \in \mathbb{R}_+$ such that for any function $w \in \mathbf{L}$ the following two conditions are satisfied:

1. for any $(x, y) \in E^2, u \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|Qw(\phi(x, u)) - Qw(\phi(y, u))| \leq [Q][w]_1 |x - y|,$$
2. for any $(x, y) \in E^2$, one has

$$|Qw(\phi(x, t^*(x))) - Qw(\phi(y, t^*(y)))| \leq [Q][w]_* |x - y|.$$

3. Quantization

The aim of this section is to describe the quantization procedure for a random variable and to recall some important properties that will be used in the sequel. There exists an extensive literature on quantization methods for random variables and processes. We do not pretend to present here an exhaustive panorama of these methods. However, the interested reader may for instance, consult the following works Gray and Neuhoff (1998); Pagès (1998); Pagès et al. (2004b) and references therein. Consider X an \mathbb{R}^q -valued random variable such that $\|X\|_p < \infty$ where $\|X\|_p$ denotes the

$$L_p\text{-nom of } X: \|X\|_p = \left(\mathbb{E}[|X|^p]\right)^{1/p}.$$

Let K be a fixed integer, the optimal L_p -quantization of the random variable X consists in finding the best possible L_p -approximation of X by a random vector \widehat{X} taking at most K values: $\widehat{X} \in \{x^1, \dots, x^K\}$. This procedure consists in the following two steps:

1. Find a finite weighted grid $\Gamma \subset \mathbb{R}^q$ with $\Gamma = \{x^1, \dots, x^K\}$.
2. Set $\widehat{X} = \widehat{X}^\Gamma$ where $\widehat{X}^\Gamma = p_\Gamma(X)$ with p_Γ denotes the closest neighbour projection on Γ .

The asymptotic properties of the L_p -quantization are given by the following result, see e.g. Pagès (1998).

Theorem 3.1. If $\mathbb{E}[|X|^{p+\eta}] < +\infty$ for some $\eta > 0$ then one has

$$\lim_{K \rightarrow \infty} K^{p/q} \min_{|\Gamma| \leq K} \|X - \widehat{X}^\Gamma\|_p^p = J_{p,q} \int |h|^{q/(q+p)}(u) du,$$

where the law of X is $P_X(du) = h(u)\lambda_q(du) + \nu$ with $\nu \perp \lambda_d, J_{p,d}$ a constant and λ_q the Lebesgue measure in \mathbb{R}^q .

Remark that X needs to have finite moments up to the order $p + \eta$ to ensure the above convergence. There exists a similar procedure for the optimal quantization of a Markov chain $\{X_k\}_{k \in \mathbb{N}}$. There

are two approaches to provide the quantized approximation of a Markov chain. The first one, based on the quantization at each time k of the random variable X_k is called the *marginal quantization*. The second one that enhances the preservation of the Markov property is called *Markovian quantization*. Remark that for the latter, the quantized Markov process is not homogeneous. These two methods are described in details in Pagès et al. (2004b, Section 3). In this work, we used the marginal quantization approach for simplicity reasons.

4. Impulse control problem

The formal probabilistic apparatus necessary to precisely define the impulse control problem is rather cumbersome, and will not be used in the sequel, therefore, for the sake of simplicity, we only present a rough description of the problem. The interested reader is referred to Costa and Davis (1989) for a rigorous definition.

A strategy $\mathcal{J} = (\tau_n, R_n)_{n \geq 1}$ is a sequence of non-anticipative intervention times $(\tau_n)_{n \geq 1}$ and non-anticipative E -valued random variables $(R_n)_{n \geq 1}$ on a measurable space $(\overline{\Omega}, \overline{\mathcal{F}})$. Between the intervention times τ_i and τ_{i+1} , the motion of the system is determined by the PDMP $\{X(t)\}$ starting from R_i . If an intervention takes place at $x \in E$, then the set of admissible points where the decision-maker can send the system to is denoted by $\mathbb{U} \subset E$. We suppose that the control set \mathbb{U} is finite and does not depend on x . The cardinal of the set \mathbb{U} is denoted by u :

$$\mathbb{U} = \{y^i : 1 \leq i \leq u\}.$$

The strategy \mathcal{J} induces a family of probability measures $P_x^\mathcal{J}, x \in E$, on $(\overline{\Omega}, \overline{\mathcal{F}})$. We define the class \mathbb{S} of admissible strategies as the strategies \mathcal{J} which satisfy $\tau_\infty = \infty$ $P_x^\mathcal{J}$ -a.s. for all $x \in E$.

Associated to the strategy \mathcal{J} , we define the following discounted cost for a process starting at $x \in E$

$$J^\mathcal{J}(x) = E_x^\mathcal{J} \left[\int_0^\infty e^{-\alpha s} f(Y_s) ds + \sum_{i=1}^\infty e^{-\alpha \tau_i} c(Y_{\tau_i}, Y_{\tau_i^+}) \right],$$

where $E_x^\mathcal{J}$ is the expectation with respect to $P_x^\mathcal{J}$ and $\{Y_t\}$ is the process with interventions. The function f then corresponds to the running cost and $c(x, y)$ corresponds to the intervention cost of moving the process from x to y , α is a positive discount factor. We make the following assumption on the cost functions.

Assumption 4.1. f is a positive function in \mathbf{L} .

Assumption 4.2. The function c is continuous on $\overline{E} \times \mathbb{U}$ and there exist $[c]_1 \in \mathbb{R}_+, [c]_2 \in \mathbb{R}_+$ and $[c]_* \in \mathbb{R}_+$ such that

1. for any $(x, y) \in E^2, u \in [0, t^*(x) \wedge t^*(y)]$,

$$\max_{z \in \mathbb{U}} |c(\phi(x, u), z) - c(\phi(y, u), z)| \leq [c]_1 |x - y|,$$
2. for any $x \in E$, and $(t, s) \in [0, t^*(x)]^2$,

$$\max_{z \in \mathbb{U}} |c(\phi(x, t), z) - c(\phi(x, s), z)| \leq [c]_2 |t - s|,$$
3. for any $(x, y) \in E^2$,

$$\max_{z \in \mathbb{U}} |c(\phi(x, t^*(x)), z) - c(\phi(y, t^*(y)), z)| \leq [c]_* |x - y|,$$
4. for any $(x, y) \in \overline{E} \times \mathbb{U}, 0 < c_0 \leq c(x, y) \leq C_c$,
5. for any $(x, y, z) \in \overline{E} \times \mathbb{U} \times \mathbb{U}$,

$$c(x, y) + c(y, z) \geq c(x, z).$$

The last assumption implies that the cost of taking two or more interventions instantaneously will not be lower than taking a single intervention. Finally, the value function for the discounted infinite horizon impulse control problem is defined for all x in E by

$$\mathcal{V}(x) = \inf_{\delta \in \mathcal{S}} \mathcal{J}^\delta(x).$$

Associated to this impulse control problem, we define the following operators. For $x \in E, t \geq 0, (v, w) \in \mathbf{C}^2$, set

$$\begin{aligned} F(x, t) &= \int_0^{t \wedge t^*(x)} e^{-\alpha s - \Lambda(x, s)} f(\phi(x, s)) ds, \\ H v(x, t) &= e^{-\alpha t \wedge t^*(x) - \Lambda(x, t \wedge t^*(x))} v(\phi(x, t \wedge t^*(x))) \\ &= \mathbb{E} \left[e^{-\alpha(t \wedge t^*(Z_0))} v(\phi(Z_0, t \wedge t^*(Z_0))) \right. \\ &\quad \left. \times \mathbf{1}_{\{S_1 \geq t \wedge t^*(Z_0)\}} \middle| Z_0 = x \right], \\ I w(x, t) &= \int_0^{t \wedge t^*(x)} e^{-\alpha s - \Lambda(x, s)} \lambda Q w(\phi(x, s)) ds \\ &= \mathbb{E} \left[e^{-\alpha S_1} w(Z_1) \mathbf{1}_{\{S_1 < t \wedge t^*(Z_0)\}} \middle| Z_0 = x \right]. \end{aligned}$$

Finally for notational convenience, let us introduce for $(v, w) \in \mathbf{C}^2, x \in E$ and $t \geq 0$.

$$\begin{aligned} J(v, w)(x, t) &= F(x, t) + H v(x, t) + I w(x, t), \\ K w(x) &= F(x, t^*(x)) + H Q w(x, t^*(x)) + I w(x, t^*(x)). \end{aligned}$$

It is easy to show that for all $n \in \mathbb{N}$

$$\begin{aligned} K v(x) &= \mathbb{E} \left[F(Z_n, t^*(Z_n)) + e^{-\alpha S_{n+1}} v(Z_{n+1}) \middle| Z_n = x \right], \\ J(v, w)(x, t) &= \mathbb{E} \left[F(Z_n, t) + e^{-\alpha S_{n+1}} w(Z_{n+1}) \mathbf{1}_{\{S_{n+1} < t \wedge t^*(Z_n)\}} \right. \\ &\quad \left. + e^{-\alpha t \wedge t^*(Z_n)} v(\phi(Z_n, t \wedge t^*(Z_n))) \right. \\ &\quad \left. \times \mathbf{1}_{\{S_{n+1} \geq t \wedge t^*(Z_n)\}} \middle| Z_n = x \right]. \end{aligned} \tag{1}$$

Note that these operators involve the original non controlled process $\{X(t)\}$ and only depend on the underlying Markov chain $(\Theta_n) = (Z_n, S_n)$. The equalities above are valid for all n because (Θ_n) is an homogeneous Markov chain. Finally, for $(v, w) \in \mathbf{C}^2, \varphi$ defined on \mathbb{U} and $x \in E$, set

$$\begin{aligned} M \varphi(x) &= \inf_{y \in \mathbb{U}} \{c(x, y) + \varphi(y)\}, \\ L(v, w)(x) &= \inf_{t \in \mathbb{R}_+} J(v, w)(x, t) \wedge K w(x), \\ \mathcal{L} w(x) &= L(M w, w)(x). \end{aligned}$$

As explained in Costa and Davis (1989), operator \mathcal{L} applied to w is the value function of the single-jump-or-intervention problem with cost function w and the value function \mathcal{V} can be computed by iterating \mathcal{L} . More precisely, let h be the cost associated to the no-impulse strategy:

$$h(x) = E_x \left[\int_0^\infty e^{-\alpha s} f(X_s) ds \right],$$

for all $x \in E$. Then we recall Proposition 4 of Costa and Davis (1989).

Proposition 4.3. Assume that g is in \mathbf{L} and $g \geq h$. Define $\mathcal{V}_0^g = g$ and $\mathcal{V}_{n+1}^g = \mathcal{L}(\mathcal{V}_n^g)$, for all $n \geq 0$. Then for all $x \in E$

$$\mathcal{V}(x) = \lim_{n \rightarrow \infty} \mathcal{V}_n^g(x).$$

As pointed out in Costa and Davis (1989), if one chooses exactly $g = h$, then \mathcal{V}_n^h corresponds to the value function of the impulse problem where only n jumps plus interventions are allowed, and after that, there are no further interventions.

Remark 4.4. Note that operator L is quite similar to the operator used in optimal stopping, see e.g. Costa and Davis (1988), de Saporta et al. (2010). However, the iteration procedure here does not rely on L but on \mathcal{L} . The difference between operators L and \mathcal{L} comes from the operator M that chooses optimally the next starting point. This is one of the main technical differences between approximating the value functions of an optimal stopping and impulse problems, and it makes the approximation scheme significantly more difficult, as explained in the next section.

5. Approximation of the value function

From now on, we assume that the distribution of $X(0)$ is given by δ_{x_0} for some fixed point x_0 in the state space E . We also choose a function g in \mathbf{L} satisfying $g \geq h$. Our approximation of the value function at x_0 is based on Proposition 4.3. Following the approach proposed by Costa and Davis (1989), we suppose now that we have selected a suitable index N such that $\mathcal{V}(x_0) - \mathcal{V}_N^g(x_0)$ is small enough see the example in Section 6. We turn to the approximation of $\mathcal{V}_N^g(x_0)$ which is the main object of this paper. In all generality, finding an index N such that $\mathcal{V}(x_0) - \mathcal{V}_N^g(x_0)$ is below a prescribed level is a very difficult problem to solve. However, in particular cases one can hope to be able to evaluate the distance between $\mathcal{V}(x_0)$ and $\mathcal{V}_N^g(x_0)$. As suggested by Costa and Davis (1989), a value of N can be chosen by calculating $\mathcal{V}_n^g(x_0)$ for different values of n and stopping when the difference between two consecutive values is small enough. Our results of convergence are derived for a fixed but arbitrary N .

Recall that if $v_0 = h$, then \mathcal{V}_N^h corresponds to the value function of the impulse problem where only N jumps plus interventions are allowed. This is an interesting problem to be solved in itself. For notational convenience, we will change our notation in the sequel and reverse the indices for the sequence $(\mathcal{V}_n^g)_{0 \leq n \leq N}$. Set

$$\begin{cases} v_N = g = \mathcal{V}_0^g, \\ v_n = \mathcal{L} v_{n+1} = \mathcal{V}_{N-n}^g, \text{ for all } 0 \leq n < N. \end{cases}$$

As explained in the introduction, the keystone of the approximation procedure for optimal stopping in de Saporta et al. (2010) is that the analogue of Proposition 4.3 yields a recursive construction of the random variables $v_n(Z_n)$. Unfortunately, this key and important property does not hold anymore here. Indeed, one has:

$$\begin{aligned} v_n(Z_n) &= \mathcal{L} v_{n+1}(Z_n) \\ &= \left(\inf_{t \in \mathbb{R}_+} \mathbb{E} \left[F(Z_n, t) + e^{-\alpha S_{n+1}} v_{n+1}(Z_{n+1}) \mathbf{1}_{\{S_{n+1} < t \wedge t^*(Z_n)\}} \right. \right. \\ &\quad \left. \left. + e^{-\alpha t \wedge t^*(Z_n)} M v_{n+1}(\phi(Z_n, t \wedge t^*(Z_n))) \right. \right. \\ &\quad \left. \left. \times \mathbf{1}_{\{S_{n+1} \geq t \wedge t^*(Z_n)\}} \middle| Z_n \right] \right) \\ &\quad \wedge \mathbb{E} \left[F(Z_n, t^*(Z_n)) + e^{-\alpha S_{n+1}} v_{n+1}(Z_{n+1}) \middle| Z_n \right]. \end{aligned}$$

And $M v_{n+1}(\phi(Z_n, t \wedge t^*(Z_n)))$ cannot be written as a function of $v_{n+1}(Z_{n+1})$. Hence, we have no recursive construction of the random variables $v_n(Z_n)$ and we cannot apply the same procedure that we used for optimal stopping. Thus, we propose a new procedure to evaluate $M v_{n+1}(\phi(Z_n, t \wedge t^*(Z_n)))$ separately from the main computation of the value function.

Note that for all $0 \leq n < N$, to compute $M v_{n+1}$ at any point, one actually only needs to evaluate the value functions v_{n+1} at the points of the control grid \mathbb{U} . We propose again a recursive computation based on the Markov chain (Z_n, S_n) but with a different starting point. Set $Z_0^y = y \in \mathbb{U}$ and $S_0^y = 0$. We denote by (Z_n^y, S_n^y) the Markov chain starting from this point $(y, 0)$. One clearly knows $v_N = g$ on \mathbb{U} . Now suppose we have computed all the v_n on \mathbb{U} for $k + 1 \leq n \leq N$. Therefore, all functions $M v_n$ are known everywhere. We can then propose the following recursive computation to evaluate v_k at $y \in \mathbb{U}$:

$$\begin{cases} v_N(Z_{N-k}^y) = g(Z_{N-k}^y) \\ v_{k+n}(Z_n^y) = \mathcal{L}(M v_{k+n+1}, v_{k+n+1})(Z_n^y), \\ \forall 0 \leq n \leq N - k - 1. \end{cases} \tag{2}$$

This way, one obtains $v_k(Z_0^y)$ that exactly equals $v_k(y)$. Note that, since the functions $M v_{k+n}$ are known, this provides a tractable recurrence relation on the random variables $v_{k+n}(Z_k^y)$.

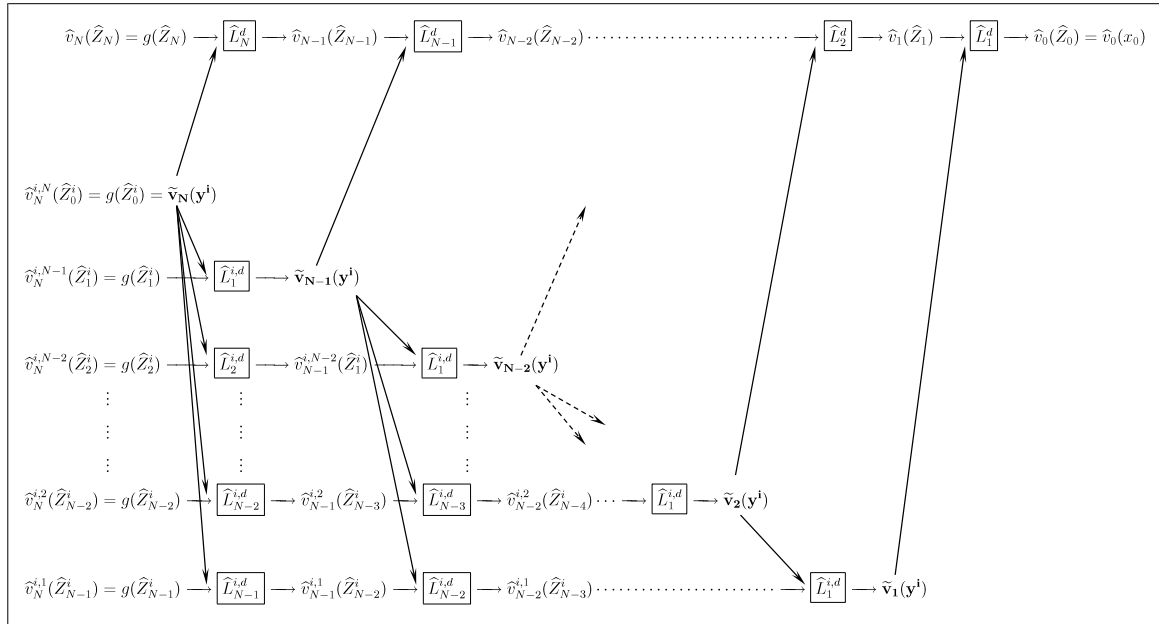


Fig. 1. Step by step procedure.

Remark 5.1. Note that this procedure requires the knowledge of function g for all the random variables $(Z_n^y)_{n \leq N-1}$ defined for the different starting points $y \in \mathbb{U}$. This is why, in general, we are not able to use the no-impulse cost function h . Indeed, it is hard to compute this function, especially if we need to know it everywhere on the state space. The most practical solution is to take g equal to a upper bound of h , and therefore constant.

There is yet another new difficulty hidden in the recurrence relation (2) above as regards its discretization. Indeed, to compute $v_n(y)$, one needs first to compute all the $v_{k+n}(Z_n^y)$ with $1 \leq n \leq N-k$, and to compute $v_{k+1}(y)$ for instance, one has already computed all the $v_{k+n}(Z_{n-1}^y)$ for $2 \leq n \leq N-k$. Unfortunately, one cannot re-use the values of $v_{k+n}(Z_{n-1}^y)$ to compute that of $v_{k+n}(Z_n^y)$, so the computation has to be started all over again each time, and one has to be very careful in the design of the approximation scheme. However, all these computations can be done with the same discretization grids for (Z_n^y, S_n^y) , so that our procedure is still reasonably fast, see Section 5.2 for details, and Fig. 1 for a graphical illustration of our procedure.

Remark 5.2. The recursive procedure (2) is triangular in the sense that one needs to compute all the $v_{k+n}(Z_n^y)$ for $0 \leq k \leq N$ and $0 \leq n \leq N-k$.

Our approximation procedure is in three steps, as explained in the following sections. The first step consists in replacing the continuous minimization in the definition of operator L by a discrete-time minimization, on path adapted grids. The second step is specific to the impulse problem, and is due to the operator M as explained in details above. The second step hence consists in carefully approximating the value functions v_n on the control grid \mathbb{U} . The last step will then be similar to the approximation of the optimal stopping problem and will consist in approximating the value functions at the points of the quantization grids of the no impulse process.

5.1. Time discretization

We define the path-adapted discretization grids as follows.

Definition 5.3. For $z \in E$, set $\Delta(z) \in]0, t^*(z)[$. Define $n(z) = \text{int}(\frac{t^*(z)}{\Delta(z)}) - 1$, where $\text{int}(x)$ denotes the greatest integer smaller

than or equal to x . The set of points $(t_i)_{i \in \{0, \dots, n(z)\}}$ with $t_i = i\Delta(z)$ is denoted by $G(z)$. This is the grid associated to the time interval $[0, t^*(z)]$.

Remark 5.4. It is important to note that, for all $z \in E$, not only one has $t^*(z) \notin G(z)$, but also $\max G(z) = t_{n(z)} \leq t^*(z) - \Delta(z)$. This property is crucial for the sequel.

We propose the following approximation of operator L , where the continuous minimization is replaced by a discrete-time minimization on the path-adapted grids.

Definition 5.5. For $(v, w) \in \mathbb{L}^2$ and $x \in E$, set

$$L^d(v, w)(x) = \min_{t \in G(x)} J(v, w)(x, t) \wedge Kw(x).$$

Now we compute the error induced by the replacement of the continuous minimization by the discrete one.

Lemma 5.6. Let $(v, w) \in \mathbb{L}^2$. Then for all $x \in E$,

$$\left| \inf_{t \leq t^*(x)} J(v, w)(x, t) - \min_{s \in G(x)} J(v, w)(x, s) \right| \leq (C_f + C_w C_\lambda + [v]_2 + C_v(C_\lambda + \alpha)) \Delta(x).$$

Proof. We have

$$\left| \inf_{t \leq t^*(x)} J(v, w)(x, t) - \min_{s \in G(x)} J(v, w)(x, s) \right| = \min_{s \in G(x)} J(v, w)(x, s) - \inf_{t \leq t^*(x)} J(v, w)(x, t).$$

Clearly, there exists $\bar{t} \in [0, t^*(x)]$ such that $\inf_{t \leq t^*(x)} J(v, w)(x, t) = J(v, w)(x, \bar{t})$. Moreover, there exists $0 \leq i \leq n(x)$ such that $\bar{t} \in [t_i, t_{i+1}[$ (with $t_{n(x)+1} = t^*(x)$). Consequently, Lemma A.5 yields

$$\left| \inf_{t \leq t^*(x)} J(v, w)(x, t) - \min_{s \in G(x)} J(v, w)(x, s) \right| \leq J(v, w)(x, t_i) - J(v, w)(x, \bar{t}) \leq (C_f + C_w C_\lambda + [v]_2 + C_v(C_\lambda + \alpha)) |\bar{t} - t_i|$$

implying the result. \square

Lemma 5.7. Let $(v, w) \in \mathbf{L}^2$ be nonnegative functions. Then for all $x \in E$,

$$|L(v, w)(x) - L^d(v, w)(x)| \leq (C_f + C_w C_\lambda + [v]_2 + C_v(C_\lambda + \alpha)) \Delta(x).$$

Proof. Since the functions v and w are nonnegative, it follows from the definition of L and L^d that

$$|L(v, w)(x) - L^d(v, w)(x)| \leq \left| \inf_{t \leq t^*(x)} J(v, w)(x, t) - \min_{s \in G(x)} J(v, w)(x, s) \right|.$$

Now in view of the previous lemma, one obtains the result. \square

5.2. Approximation of the value functions on the control grid \mathbb{U}

We now need to introduce the quantized approximations of the underlying Markov chains (Θ_n^y) . More precisely, we need several approximations at this stage, one for each starting point y in the control set \mathbb{U} . Recall that $\mathbb{U} = \{y^i, 1 \leq i \leq u\}$. For all $1 \leq i \leq u$, let $(Z_n^i, S_n^i)_{0 \leq n \leq N-1}$ be the Markov chain $(Z_n, S_n)_{0 \leq n \leq N-1}$ with starting point $Z_0 = y^i, S_0 = 0$, and let $(\widehat{Z}_n^i, \widehat{S}_n^i)_{0 \leq n \leq N-1}$ be the quantized approximation of the sequence $(Z_n^i, S_n^i)_{0 \leq n \leq N-1}$, see Section 3. The quantization algorithm provides us with a finite grid $\Gamma_n^{i,\Theta} \subset E \times \mathbb{R}_+$ at each time $0 \leq n \leq N - 1$ as well as weights for each point of the grid and transition probabilities from one grid to the next one, see e.g. Bally and Pagès (2003), Pagès (1998), Pagès et al. (2004b) for details. Set $p \geq 1$ such that Θ_n has finite moments at least up to the order $p + \epsilon$ for some positive ϵ and let p_n^i be the closest-neighbour projection from $E \times \mathbb{R}_+$ onto $\Gamma_n^{i,\Theta}$ (for the distance of norm p ; if there are several equally close neighbours, pick the one with the smallest index). Then the quantization of Θ_n^i conditionally to $Z_0 = y^i$ is defined by

$$\widehat{\Theta}_n^i = (\widehat{Z}_n^i, \widehat{S}_n^i) = p_n^i(Z_n^i, S_n^i).$$

We will also denote $\Gamma_n^{i,Z}$ the projection of $\Gamma_n^{i,\Theta}$ on E and $\Gamma_n^{i,S}$ the projection of $\Gamma_n^{i,\Theta}$ on \mathbb{R}_+ .

Although (Z_n^i, S_n^i) is a Markov chain, its quantized approximation is usually not a Markov chain. It can be turned into a Markov chain by slightly changing the ponderations in the grids, see Pagès et al. (2004a), but this Markov chain will not be homogeneous in any case. Therefore, the following quantized approximations of operators H, I, K, J and L^d depend on both indices n and i .

Definition 5.8. For $v \in \mathbf{L}^2, w$ defined on $\Gamma_{n+1}^{i,Z}, x \in E, 0 \leq n \leq N - 1, 1 \leq i \leq u$ and $z \in \Gamma_n^{i,Z}$, consider

$$\begin{aligned} \widehat{H}_{n+1}^i v(z, t) &= \mathbb{E} \left[e^{-\alpha(t \wedge t^*(\widehat{Z}_n^i))} v(\phi(\widehat{Z}_n^i, t \wedge t^*(\widehat{Z}_n^i))) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1}^i \geq t \wedge t^*(\widehat{Z}_n^i)\}} \Big| \widehat{Z}_n^i = z \right], \\ \widehat{I}_{n+1}^i w(z, t) &= \mathbb{E} \left[e^{-\alpha \widehat{S}_{n+1}^i} w(\widehat{Z}_{n+1}^i) \mathbf{1}_{\{\widehat{S}_{n+1}^i < t \wedge t^*(\widehat{Z}_n^i)\}} \Big| \widehat{Z}_n^i = z \right], \\ \widehat{K}_{n+1}^i w(z) &= \mathbb{E} \left[F(\widehat{Z}_n^i, t^*(\widehat{Z}_n^i)) + e^{-\alpha \widehat{S}_{n+1}^i} w(\widehat{Z}_{n+1}^i) \Big| \widehat{Z}_n^i = z \right], \end{aligned}$$

$$\begin{aligned} \widehat{J}_{n+1}^i(v, w)(z, t) &= \mathbb{E} \left[F(\widehat{Z}_n^i, t) + e^{-\alpha \widehat{S}_{n+1}^i} w(\widehat{Z}_{n+1}^i) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1}^i < t \wedge t^*(\widehat{Z}_n^i)\}} \Big| \widehat{Z}_n^i = z \right] \\ &\quad + \mathbb{E} \left[e^{-\alpha(t \wedge t^*(\widehat{Z}_n^i))} v(\phi(\widehat{Z}_n^i, t \wedge t^*(\widehat{Z}_n^i))) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1}^i \geq t \wedge t^*(\widehat{Z}_n^i)\}} \Big| \widehat{Z}_n^i = z \right], \\ \widehat{I}_{n+1}^{i,d}(v, w)(z) &= \min_{t \in G(z)} \widehat{J}_{n+1}^i(v, w)(z, t) \wedge \widehat{K}_{n+1}^i w(z). \end{aligned}$$

Our approximation scheme goes backwards in time, in as much as it is initialized with computing v_N at the points of the last quantization grids $\Gamma_N^{i,Z}$, then v_{N-1} is computed on $\Gamma_{N-1}^{i,Z}$ and so on.

Definition 5.9. Set $\widetilde{v}_N(y^i) = g(y^i)$ for $1 \leq i \leq u$. Then, for $1 \leq k \leq N - 1$ and $1 \leq i \leq u$, set $\widetilde{v}_k(y^i) = \widehat{v}_k^{i,k}(y^i)$, where

$$\begin{aligned} \widehat{v}_N^{i,k}(z) &= g(z), \quad z \in \Gamma_{N-k}^{i,Z}, \\ \widehat{v}_{k+n-1}^{i,k}(z) &= \widehat{I}_n^{i,d}(M \widetilde{v}_{k+n}, \widehat{v}_{k+n}^k)(z), \\ &\quad z \in \Gamma_{n-1}^{i,Z}, n \in \{1, \dots, N - k\}. \end{aligned}$$

See Fig. 1 for a graphical illustration of this numerical procedure.

Remark 5.10. Note the use of both \widetilde{v}_{k+n} and \widehat{v}_{k+n}^k in the scheme above. This is due to the fact that we have to reset all our calculations for each value function \widetilde{v}_k and cannot use the calculations made for e.g. \widetilde{v}_{k+1} because the value functions are evaluated at different points, and are approximated with different discrete operators. This is mostly because the quantized process $(\widehat{Z}_n^i, \widehat{S}_n^i)$ is not an homogeneous Markov chain.

We can now state our first result on the convergence rate of this approximation.

Theorem 5.11. For all $1 \leq k \leq N - 1, 0 \leq n \leq N - k - 1$ and $1 \leq i \leq d$, suppose that $\Delta(z)$ for $z \in \Gamma_n^{i,Z}$ is such that

$$\sqrt{\frac{d^4 \|Z_n^i - \widehat{Z}_n^i\|_p + d^5 \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p}{d^3}} < \min_{z \in \Gamma_n^{i,Z}} \{\Delta(z)\}.$$

Then we have

$$\begin{aligned} &\|v_{k+n}(Z_n^i) - \widehat{v}_{k+n}^{i,k}(\widehat{Z}_n^i)\|_p \\ &\leq \|v_{k+n+1}(Z_{n+1}^i) - \widehat{v}_{k+n+1}^{i,k}(\widehat{Z}_{n+1}^i)\|_p \\ &\quad + \max_{y \in \mathbb{U}} |v_{k+n+1}(y) - \widetilde{v}_{k+n+1}(y)| \\ &\quad + d_{k,n}^1 \|Z_n^i - \widehat{Z}_n^i\|_p + 2[v_{k+n+1}] \|Z_{n+1}^i - \widehat{Z}_{n+1}^i\|_p \\ &\quad + C_f \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p + d_{k,n}^2 \|\Delta(\widehat{Z}_n^i)\|_p \\ &\quad + 2\sqrt{d^3 (d^4 \|Z_n^i - \widehat{Z}_n^i\|_p + d^5 \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p)}, \end{aligned}$$

with

$$\begin{aligned} d_{k,n}^1 &= \left\{ [Q][v_{k+n+1}]_* + 2E_3 \right\} \vee \left\{ C_c (E_1 + \alpha[t^*]) + 2([c]_1 \right. \\ &\quad \left. + [c]_2[t^*]) \right\} + [v_{k+n}] + [Q][v_{k+n+1}]_1 \frac{C_\lambda}{\alpha} + \frac{C_f}{\alpha} (E_1 + E_2), \end{aligned}$$

$$\begin{aligned}
 d_{k,n}^2 &= C_f + C_{v_{k+n+1}} C_\lambda + [c]_2 + (C_c + C_{v_{k+n+1}})(C_\lambda + \alpha), \\
 d^3 &= \left(\frac{2C_f}{\alpha} + C_c \right) C_\lambda, \\
 d^4 &= \frac{C_f}{\alpha} (1 + [t^*]) + C_c [t^*], \\
 d^5 &= 2 \left(2 \frac{C_f}{\alpha} + C_c \right).
 \end{aligned}$$

Remark 5.12. Recall that $v_N = \widehat{v}_N^{i,k} = \widetilde{v}_N = g$. Hence, one has $\|v_N(Z_{N-k}^i) - \widehat{v}_N^{i,k}(\widehat{Z}_{N-k}^i)\|_p \leq [g] \|\widehat{Z}_N^i - Z_N^i\|_p$ and $\max_{y \in \mathbb{U}} |v_N(y) - \widetilde{v}_N(y)| = 0$. In addition, the quantization error $\|\phi_n^i - \widehat{\phi}_n^i\|_p$ goes to zero as the number of points in the grids goes to infinity, see e.g. Pagès (1998). Therefore, according to Definition 5.9 and by using an induction procedure $\max_{y \in \mathbb{U}} |v_k(y) - \widetilde{v}_k(y)|$ can be made arbitrarily small by an adequate choice of the discretization parameters. From a theoretical point of view, the error can be calculated by iterating the result of Theorem 5.11. However, this result is not presented here because it would lead to an intricate expression. From a numerical point of view, a computer can easily estimate this error as shown in the example of Section 6.

The proof is going to be detailed in the following sections. We first split the error into four terms. For all $1 \leq k \leq N - 1$, $0 \leq n \leq N - k - 1$ and $1 \leq i \leq d$, we have

$$\|v_{k+n}(Z_n^i) - \widehat{v}_{k+n}^{i,k}(\widehat{Z}_n^i)\|_p \leq \sum_{j=1}^4 \mathcal{Y}_j^i,$$

where

$$\begin{aligned}
 \mathcal{Y}_1^i &= \|v_{k+n}(Z_n^i) - v_{k+n}(\widehat{Z}_n^i)\|_p, \\
 \mathcal{Y}_2^i &= \|L(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - L^d(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i)\|_p, \\
 \mathcal{Y}_3^i &= \|L^d(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - \widehat{L}_{n+1}^{i,d}(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i)\|_p, \\
 \mathcal{Y}_4^i &= \left\| \widehat{L}_{n+1}^{i,d}(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - \widehat{L}_{n+1}^{i,d}(M\widetilde{v}_{k+n+1}, \widehat{v}_{k+n+1}^{i,k})(\widehat{Z}_n^i) \right\|_p.
 \end{aligned}$$

The first two terms are easy enough to handle thanks to Corollary A.12 and Lemma 5.7.

Lemma 5.13. A upper bound for \mathcal{Y}_1^i is

$$\|v_{k+n}(Z_n^i) - v_{k+n}(\widehat{Z}_n^i)\|_p \leq [v_{k+n}] \|Z_n^i - \widehat{Z}_n^i\|_p.$$

Lemma 5.14. A upper bound for \mathcal{Y}_2^i is

$$\begin{aligned}
 &\|L(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - L^d(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i)\|_p \\
 &\leq \left(C_f + C_{v_{k+n+1}} C_\lambda + [c]_2 + (C_c + C_{v_{k+n+1}})(C_\lambda + \alpha) \right) \|\Delta(\widehat{Z}_n^i)\|_p.
 \end{aligned}$$

The fourth term is also easy enough to deal with as it is a mere comparison of two finite weighted sums.

Lemma 5.15. A upper bound for \mathcal{Y}_4^i is

$$\begin{aligned}
 &\left\| \widehat{L}_{n+1}^{i,d}(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - \widehat{L}_{n+1}^{i,d}(M\widetilde{v}_{k+n+1}, \widehat{v}_{k+n+1}^{i,k})(\widehat{Z}_n^i) \right\|_p
 \end{aligned}$$

$$\begin{aligned}
 &\leq [v_{k+n+1}] \|Z_{n+1}^i - \widehat{Z}_{n+1}^i\|_p \\
 &\quad + \left\| v_{k+n+1}(Z_{n+1}^i) - \widehat{v}_{k+n+1}^{i,k}(\widehat{Z}_{n+1}^i) \right\|_p \\
 &\quad + \max_{y \in \mathbb{U}} |v_{k+n+1}(y) - \widetilde{v}_{k+n+1}(y)|.
 \end{aligned}$$

Proof. We clearly have

$$\begin{aligned}
 &\left\| \widehat{L}_{n+1}^{i,d}(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) - \widehat{L}_{n+1}^{i,d}(M\widetilde{v}_{k+n+1}, \widehat{v}_{k+n+1}^{i,k})(\widehat{Z}_n^i) \right\|_p \\
 &\leq \left\| \max_{t \in G(\widehat{Z}_n^i)} \widehat{J}_{n+1}^i(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i, t) - \widehat{J}_{n+1}^i(M\widetilde{v}_{k+n+1}, \widehat{v}_{k+n+1}^{i,k})(\widehat{Z}_n^i, t) \right\|_p \\
 &\quad \vee \left\| \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) - \widehat{K}_{n+1}^i \widehat{v}_{k+n+1}^{i,k}(\widehat{Z}_n^i) \right\|_p \\
 &\leq \left\| \mathbb{E}[v_{k+n+1}(\widehat{Z}_{n+1}^i) - \widehat{v}_{k+n+1}^{i,k}(\widehat{Z}_{n+1}^i) | \widehat{Z}_n^i] \right\|_p \\
 &\quad + \left\| \mathbb{E}\left[Mv_{k+n+1}(\phi(\widehat{Z}_n^i, t \wedge t^*(\widehat{Z}_n^i))) - M\widetilde{v}_{k+n+1}(\phi(\widehat{Z}_n^i, t \wedge t^*(\widehat{Z}_n^i))) \mid \widehat{Z}_n^i \right] \right\|_p \\
 &\leq \|v_{k+n+1}(\widehat{Z}_{n+1}^i) - v_{k+n+1}(Z_{n+1}^i)\|_p \\
 &\quad + \left\| v_{k+n+1}(Z_{n+1}^i) - \widehat{v}_{k+n+1}^{i,k}(\widehat{Z}_{n+1}^i) \right\|_p \\
 &\quad + \max_{y \in \mathbb{U}} |v_{k+n+1}(y) - \widetilde{v}_{k+n+1}(y)|,
 \end{aligned}$$

showing the result. \square

We now turn to the third term. This is the key step of the error evaluation, because on the one hand, this is where we deal with the indicator functions. The main idea is that although they are not continuous, we prove in the following two lemmas that the set where the discontinuity actually occurs is of small enough probability. This is also where our special choice of time discretization grids is crucial. On the other hand, we use here the specific properties of quantization.

Lemma 5.16. For all $1 \leq i \leq d$, $n \in \{0, \dots, N - 1\}$ and $0 < \eta < \min_{z \in \Gamma_n^{i,z}} \{\Delta(z)\}$,

$$\left\| \mathbf{1}_{t^*(Z_n^i) < t^*(\widehat{Z}_n^i) - \eta} \right\|_p \leq \frac{[t^*] \|Z_n^i - \widehat{Z}_n^i\|_p}{\eta}.$$

Proof. By using the Chebyshev's inequality, one clearly has

$$\begin{aligned}
 \mathbb{E} \left[\left| \mathbf{1}_{t^*(Z_n^i) < t^*(\widehat{Z}_n^i) - \eta} \right|^p \right] &= \mathbf{P}(t^*(Z_n^i) < t^*(\widehat{Z}_n^i) - \eta) \\
 &\leq \mathbf{P}(|t^*(Z_n^i) - t^*(\widehat{Z}_n^i)| > \eta) \\
 &\leq \frac{[t^*]^p \|Z_n^i - \widehat{Z}_n^i\|_p^p}{\eta^p},
 \end{aligned}$$

showing the result. \square

Lemma 5.17. For all $1 \leq i \leq d$, $n \in \{0, \dots, N - 1\}$ and $0 < \eta < \min_{z \in \Gamma_n^{i,z}} \{\Delta(z)\}$,

$$\begin{aligned}
 &\left\| \max_{s \in G(\widehat{Z}_n^i)} \mathbb{E} \left[\left| \mathbf{1}_{S_{n+1}^i < s \wedge t^*(Z_n^i)} - \mathbf{1}_{\widehat{S}_{n+1}^i < s \wedge t^*(\widehat{Z}_n^i)} \right| \mid \widehat{Z}_n^i \right] \right\|_p \\
 &\leq \frac{2}{\eta} \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p + C_\lambda \eta + \frac{2[t^*]}{\eta} \|Z_n^i - \widehat{Z}_n^i\|_p.
 \end{aligned}$$

Proof. Set $0 < \eta < \min_{z \in \Gamma_n^{i,z}} \{\Delta(z)\}$ and $s \in G(\widehat{Z}_n^i)$. By definition of the grid $G(\widehat{Z}_n^i)$ and η , one has $s + \eta < t^*(\widehat{Z}_n^i)$, see Remark 5.4. Thus, the difference of indicator functions can be written as

$$\begin{aligned} & \left| \mathbf{1}_{\{S_{n+1}^i < s \wedge t^*(Z_n^i)\}} - \mathbf{1}_{\{\widehat{S}_{n+1}^i < s \wedge t^*(\widehat{Z}_n^i)\}} \right| \\ & \leq \left| \mathbf{1}_{\{S_{n+1}^i < s \wedge t^*(Z_n^i)\}} - \mathbf{1}_{\{\widehat{S}_{n+1}^i < s \wedge t^*(\widehat{Z}_n^i)\}} \right| \left[\mathbf{1}_{\{t^*(Z_n^i) \leq t^*(\widehat{Z}_n^i) - \frac{\eta}{2}\}} \right. \\ & \quad \left. + \mathbf{1}_{\{t^*(Z_n^i) > t^*(\widehat{Z}_n^i) - \frac{\eta}{2}\}} \right] \\ & \leq \mathbf{1}_{\{t^*(Z_n^i) \leq t^*(\widehat{Z}_n^i) - \frac{\eta}{2}\}} + \mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \left| \mathbf{1}_{\{S_{n+1}^i < s\}} - \mathbf{1}_{\{\widehat{S}_{n+1}^i < s\}} \right| \\ & \leq \mathbf{1}_{\{t^*(Z_n^i) \leq t^*(\widehat{Z}_n^i) - \frac{\eta}{2}\}} + \mathbf{1}_{\{|S_{n+1}^i - \widehat{S}_{n+1}^i| > \frac{\eta}{2}\}} \\ & \quad + \mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \mathbf{1}_{\{|S_{n+1}^i - s| \leq \frac{\eta}{2}\}}. \end{aligned}$$

This yields

$$\begin{aligned} & \left\| \max_{s \in G(\widehat{Z}_n^i)} \mathbb{E} \left[\left| \mathbf{1}_{\{S_{n+1}^i < s \wedge t^*(Z_n^i)\}} - \mathbf{1}_{\{\widehat{S}_{n+1}^i < s \wedge t^*(\widehat{Z}_n^i)\}} \right| \middle| \widehat{Z}_n^i \right] \right\|_p \\ & \leq \left\| \mathbf{1}_{\{t^*(Z_n^i) \leq t^*(\widehat{Z}_n^i) - \frac{\eta}{2}\}} \right\|_p + \left\| \mathbf{1}_{\{|S_{n+1}^i - \widehat{S}_{n+1}^i| > \frac{\eta}{2}\}} \right\|_p \\ & \quad + \left\| \max_{s \in G(\widehat{Z}_n^i)} \mathbb{E} \left[\mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \mathbf{1}_{\{|S_{n+1}^i - s| \leq \frac{\eta}{2}\}} \right] \middle| \widehat{Z}_n^i \right] \right\|_p. \end{aligned} \tag{3}$$

On the one hand, Chebyshev's inequality gives

$$\begin{aligned} \left\| \mathbf{1}_{\{|S_{n+1}^i - \widehat{S}_{n+1}^i| > \frac{\eta}{2}\}} \right\|_p &= \mathbf{P} \left(|S_{n+1}^i - \widehat{S}_{n+1}^i| > \frac{\eta}{2} \right) \\ &\leq \frac{2^p \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p^p}{\eta^p}. \end{aligned} \tag{4}$$

On the other hand, one has

$$\begin{aligned} & \mathbb{E} \left[\mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \mathbf{1}_{\{|S_{n+1}^i - s| \leq \frac{\eta}{2}\}} \middle| \widehat{Z}_n^i \right] \\ &= \mathbb{E} \left[\mathbb{E} \left[\mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \mathbf{1}_{\{s - \frac{\eta}{2} \leq S_{n+1}^i \leq s + \frac{\eta}{2}\}} \middle| Z_n^i \right] \middle| \widehat{Z}_n^i \right] \\ &= \mathbb{E} \left[\mathbf{1}_{\{t^*(Z_n^i) > s + \frac{\eta}{2}\}} \int_{s - \frac{\eta}{2}}^{s + \frac{\eta}{2}} \lambda(\phi(Z_n^i, u)) du \middle| \widehat{Z}_n^i \right] \\ &\leq \eta C_\lambda. \end{aligned} \tag{5}$$

Combining Lemma 5.16 and Eqs. (3)–(5), the result follows. \square

We now look up the error made in replacing K by \widehat{K}_{n+1}^i . This is where we use the specific properties of quantization.

Lemma 5.18. For all $1 \leq i \leq d$, $k \in \{1, \dots, N - 1\}$ and $n \in \{1, \dots, N - k\}$, one has

$$\begin{aligned} & \left\| K v_{k+n+1}(\widehat{Z}_n^i) - \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) \right\|_p \\ & \leq C_f \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p + [v_{k+n+1}] \|Z_{n+1}^i - \widehat{Z}_{n+1}^i\|_p \\ & \quad + \left\{ [Q][v_{k+n+1}]_1 \frac{C_\lambda}{\alpha} + [Q][v_{k+n+1}]_* + \frac{C_f}{\alpha} (E_1 + E_2) \right. \\ & \quad \left. + 2E_3 \right\} \|Z_n^i - \widehat{Z}_n^i\|_p. \end{aligned}$$

Proof. We have

$$\begin{aligned} & \left| K v_{k+n+1}(\widehat{Z}_n^i) - \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) \right| \\ & \leq \left| K v_{k+n+1}(\widehat{Z}_n^i) - \mathbb{E}[K v_{k+n+1}(Z_n^i) | \widehat{Z}_n^i] \right| \\ & \quad + \left| \mathbb{E}[K v_{k+n+1}(Z_n^i) | \widehat{Z}_n^i] - \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) \right| \\ & \leq \mathbb{E} \left[\left| K v_{k+n+1}(\widehat{Z}_n^i) - K v_{k+n+1}(Z_n^i) \right| \middle| \widehat{Z}_n^i \right] \\ & \quad + \left| \mathbb{E}[K v_{k+n+1}(Z_n^i) | \widehat{Z}_n^i] - \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) \right|. \end{aligned} \tag{6}$$

By using the Lipschitz property of K stated in Lemma A.4, we obtain

$$\begin{aligned} & \left\| \mathbb{E} \left[\left| K v_{k+n+1}(\widehat{Z}_n^i) - K v_{k+n+1}(Z_n^i) \right| \middle| \widehat{Z}_n^i \right] \right\|_p \\ & \leq \left\{ [Q][v_{k+n+1}]_1 \frac{C_\lambda}{\alpha} + [Q][v_{k+n+1}]_* \right. \\ & \quad \left. + C_{v_{k+n+1}} (E_1 + E_2) + E_3 \right\} \|Z_n^i - \widehat{Z}_n^i\|_p. \end{aligned} \tag{7}$$

Then, recall that by construction of the quantized process, one has $(\widehat{Z}_n^i, \widehat{S}_n^i) = p_n^i(Z_n^i, S_n^i)$. Hence we have the following crucial property: $\sigma\{\widehat{Z}_n^i\} \subset \sigma\{Z_n^i, S_n^i\}$. By using the special structure of the PDMP $\{X(t)\}$, we also have $\sigma\{Z_n^i, S_n^i\} \subset \mathcal{F}_{T_n}$, so that one has $\sigma\{\widehat{Z}_n^i\} \subset \sigma\{Z_n^i\}$. It now follows from the definition of K given in Eq. (1) that

$$\begin{aligned} & \left| \mathbb{E}[K v_{k+n+1}(Z_n^i) | \widehat{Z}_n^i] - \widehat{K}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i) \right| \\ & \leq \mathbb{E} \left[\left| F(Z_n^i, t^*(Z_n^i)) - F(\widehat{Z}_n^i, t^*(\widehat{Z}_n^i)) \right| \middle| \widehat{Z}_n^i \right] \\ & \quad + \mathbb{E} \left[\left| e^{-\alpha S_{n+1}^i} v_{k+n+1}(Z_{n+1}^i) - e^{-\alpha \widehat{S}_{n+1}^i} v_{k+n+1}(\widehat{Z}_{n+1}^i) \right| \middle| \widehat{Z}_n^i \right]. \end{aligned} \tag{8}$$

From Lemma A.3, we readily obtain

$$\begin{aligned} & \left\| \mathbb{E} \left[\left| F(Z_n^i, t^*(Z_n^i)) - F(\widehat{Z}_n^i, t^*(\widehat{Z}_n^i)) \right| \middle| \widehat{Z}_n^i \right] \right\|_p \\ & \leq E_3 \|Z_n^i - \widehat{Z}_n^i\|_p, \end{aligned} \tag{9}$$

and it is easy to show that

$$\begin{aligned} & \left\| \mathbb{E} \left[\left| e^{-\alpha S_{n+1}^i} v_{k+n+1}(Z_{n+1}^i) - e^{-\alpha \widehat{S}_{n+1}^i} v_{k+n+1}(\widehat{Z}_{n+1}^i) \right| \middle| \widehat{Z}_n^i \right] \right\|_p \\ & \leq [v_{k+n+1}] \|Z_{n+1}^i - \widehat{Z}_{n+1}^i\|_p + \alpha C_{v_{k+n+1}} \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p. \end{aligned} \tag{10}$$

Finally, recalling that $C_{v_{k+n}} \leq \frac{C_f}{\alpha}$ and combining Eqs. (6)–(10) we obtain the expected result. \square

We turn to the error made in replacing J by \widehat{J}_{n+1}^i . Here we use the specific properties of quantization again, and the lemmas on indicator functions.

Lemma 5.19. For all $1 \leq i \leq d$, $k \in \{1, \dots, N - 1\}$, $n \in \{1, \dots, N - k\}$, and $0 < \eta < \min_{z \in \Gamma_n^{i,z}} \{\Delta(z)\}$, one has

$$\begin{aligned} & \left\| \max_{t \in G(\widehat{Z}_n^i)} \left| J(M v_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i, t) \right. \right. \\ & \quad \left. \left. - \widehat{J}_{n+1}^i(M v_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i, t) \right| \right\|_p \\ & \leq \left\{ [Q][v_{k+n+1}]_1 \frac{C_\lambda}{\alpha} + \frac{C_f}{\alpha} (E_1 + E_2 + \alpha[t^*]) \right\} \end{aligned}$$

$$\begin{aligned}
 &+ C_c (E_1 + \alpha [t^*]) \\
 &+ 2 ([c]_1 + [c]_2 [t^*]) + \frac{1}{\eta} \left(\frac{C_f}{\alpha} (1 + [t^*]) \right. \\
 &\left. + C_c [t^*] \right) \left\| Z_n^i - \widehat{Z}_n^i \right\|_p \\
 &+ \left\{ \frac{2}{\eta} \left(2 \frac{C_f}{\alpha} + C_c \right) + C_f \right\} \left\| S_{n+1}^i - \widehat{S}_{n+1}^i \right\|_p \\
 &+ [v_{k+n+1}] \left\| Z_{n+1}^i - \widehat{Z}_{n+1}^i \right\|_p + \left(\frac{2C_f}{\alpha} + C_c \right) C_\lambda \eta.
 \end{aligned}$$

Proof. By definition of J , we have

$$\begin{aligned}
 &|J(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i, t) \\
 &\quad - \widehat{J}_{n+1}^i(Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i, t)| \\
 &\leq |Iv_{k+n+1}(\widehat{Z}_n^i, t) - \widehat{I}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i, t)| \\
 &\quad + |HMv_{k+n+1}(\widehat{Z}_n^i, t) - \widehat{H}_{n+1}^i Mv_{k+n+1}(\widehat{Z}_n^i, t)|. \tag{11}
 \end{aligned}$$

For the first term on the right hand side of Eq. (11), we proceed as for K in the preceding lemma

$$\begin{aligned}
 &|Iv_{k+n+1}(\widehat{Z}_n^i, t) - \widehat{I}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i, t)| \\
 &\leq \mathbb{E} \left[|Iv_{k+n+1}(\widehat{Z}_n^i, t) - Iv_{k+n+1}(Z_n^i, t)| \middle| \widehat{Z}_n^i \right] \\
 &\quad + \left| \mathbb{E} [Iv_{k+n+1}(Z_n^i, t) | \widehat{Z}_n^i] - \widehat{I}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i, t) \right|.
 \end{aligned}$$

On the one hand, it follows from Lemma A.3 that

$$\begin{aligned}
 &\left\| \max_{t \in G(\widehat{Z}_n^i)} \mathbb{E} \left[|Iv_{k+n+1}(\widehat{Z}_n^i, t) - Iv_{k+n+1}(Z_n^i, t)| \middle| \widehat{Z}_n^i \right] \right\|_p \\
 &\leq \left\{ \frac{1}{\alpha} ([Q][v_{k+n+1}]_1 C_\lambda + C_{v_{k+n+1}} [\lambda]_1 (1 + C_\lambda C_{\lambda^*})) \right. \\
 &\quad \left. + C_{v_{k+n+1}} C_\lambda [t^*] \right\} \left\| Z_n^i - \widehat{Z}_n^i \right\|_p.
 \end{aligned}$$

On the other hand, we use again the fact that $\sigma\{\widehat{Z}_n^i\} \subset \sigma\{Z_n^i\}$ to obtain

$$\begin{aligned}
 &|\mathbb{E} [Iv_{k+n+1}(Z_n^i, t) | \widehat{Z}_n^i] - \widehat{I}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i, t)| \\
 &\leq \mathbb{E} \left[\mathbf{1}_{\{S_{n+1}^i < t \wedge t^*(Z_n^i)\}} | e^{-\alpha S_{n+1}^i} v_{k+n+1}(Z_n^i) \right. \\
 &\quad \left. - e^{-\alpha \widehat{S}_{n+1}^i} v_{k+n+1}(\widehat{Z}_n^i) \right] \left| \widehat{Z}_n^i \right| \\
 &+ \mathbb{E} \left[e^{-\alpha \widehat{S}_{n+1}^i} v_{k+n+1}(\widehat{Z}_{n+1}^i) | \mathbf{1}_{\{S_{n+1}^i < t \wedge t^*(Z_n^i)\}} \right. \\
 &\quad \left. - \mathbf{1}_{\{\widehat{S}_{n+1}^i < t \wedge t^*(\widehat{Z}_n^i)\}} \right] \left| \widehat{Z}_n^i \right|.
 \end{aligned}$$

It remains to deal with the indicator function. Lemma 5.17 yields

$$\begin{aligned}
 &\left\| \max_{t \in G(\widehat{Z}_n^i)} |\mathbb{E} [Iv_{k+n+1}(Z_n^i, t) | \widehat{Z}_n^i] - \widehat{I}_{n+1}^i v_{k+n+1}(\widehat{Z}_n^i, t)| \right\|_p \\
 &\leq C_{v_{k+n+1}} C_\lambda \eta + \left(\alpha + \frac{2}{\eta} \right) C_{v_{k+n+1}} \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p \\
 &\quad + [v_{k+n+1}] \left\| Z_{n+1}^i - \widehat{Z}_{n+1}^i \right\|_p + \frac{2C_{v_{k+n+1}} [t^*]}{\eta} \left\| Z_n^i - \widehat{Z}_n^i \right\|_p.
 \end{aligned}$$

By using the same arguments and Lemmas A.1 and A.3, we obtain similar results for the second term on the right hand side of Eq. (11), namely

$$\begin{aligned}
 &\left\| \max_{t \in G(\widehat{Z}_n^i)} \mathbb{E} \left[|HMv_{k+n+1}(\widehat{Z}_n^i, t) - HMv_{k+n+1}(Z_n^i, t)| \middle| \widehat{Z}_n^i \right] \right\|_p \\
 &\leq \{ [c]_1 + [c]_2 [t^*] + (C_{v_{k+n+1}} + C_c) (C_{t^*} [\lambda]_1 \\
 &\quad + (C_\lambda + \alpha) [t^*]) \} \left\| Z_n^i - \widehat{Z}_n^i \right\|_p,
 \end{aligned}$$

and

$$\begin{aligned}
 &\left\| \max_{t \in G(\widehat{Z}_n^i)} |\mathbb{E} [HMv_{k+n+1}(Z_n^i, t) | \widehat{Z}_n^i] - \widehat{H}_{n+1}^i Mv_{k+n+1}(\widehat{Z}_n^i, t)| \right\|_p \\
 &\leq \frac{2}{\eta} (C_{v_{k+n+1}} + C_c) \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p + C_\lambda (C_{v_{k+n+1}} + C_c) \eta \\
 &\quad + \left\{ \alpha [t^*] (C_{v_{k+n+1}} + C_c) + [c]_1 + [c]_2 [t^*] \right. \\
 &\quad \left. + \frac{2[t^*]}{\eta} (C_{v_{k+n+1}} + C_c) \right\} \left\| Z_n^i - \widehat{Z}_n^i \right\|_p,
 \end{aligned}$$

showing the result. \square

We now add up the preceding results to obtain the following upper bound for \mathcal{Y}_3^i .

Lemma 5.20. A upper bound for \mathcal{Y}_3^i is

$$\begin{aligned}
 &\left\| L^d (Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) \right. \\
 &\quad \left. - \widehat{L}_{n+1}^{i,d} (Mv_{k+n+1}, v_{k+n+1})(\widehat{Z}_n^i) \right\|_p \\
 &\leq \|Z_n^i - \widehat{Z}_n^i\|_p \left\{ [Q][v_{k+n+1}]_1 \frac{C_\lambda}{\alpha} + \frac{C_f}{\alpha} (E_1 + E_2) \right. \\
 &\quad + \{ [Q][v_{k+n+1}]_* + 2E_3 \} \vee \left\{ C_c (E_1 + \alpha [t^*]) \right. \\
 &\quad \left. + 2 ([c]_1 + [c]_2 [t^*]) + \frac{1}{\eta} \left(\frac{C_f}{\alpha} (1 + [t^*]) + C_c [t^*] \right) \right\} \left\| \right. \\
 &\quad \left. + \|S_{n+1}^i - \widehat{S}_{n+1}^i\|_p \left\{ C_f + \frac{2}{\eta} \left(2 \frac{C_f}{\alpha} + C_c \right) \right\} \right. \\
 &\quad \left. + [v_{k+n+1}] \left\| Z_{n+1}^i - \widehat{Z}_{n+1}^i \right\|_p + \left(\frac{2C_f}{\alpha} + C_c \right) C_\lambda \eta. \right.
 \end{aligned}$$

5.3. Approximation of the value function

Now we have computed the value functions on the control grid, we turn to the actual approximation of v_0 . As in the preceding section, we define the quantized approximation of the underlying Markov chain (Θ_n) starting from $(x_0, 0)$, the actual starting point of the PDMP. Let $(Z_n, S_n)_{0 \leq n \leq N-1}$ be the quantized approximation of the sequence $(Z, S)_{0 \leq n \leq N-1}$. The quantization algorithm provides us with another series of finite grids $\Gamma_n^\Theta \subset E \times \mathbb{R}_+$ for all $0 \leq n \leq N - 1$ as well as weights for each point of the grids and transition probabilities from one grid to the next one. Let p_n be the closest-neighbour projection from $E \times \mathbb{R}_+$ onto Γ_n^Θ . Then the quantization of Θ_n conditionally to $Z_0 = x_0$ is defined by

$$\widehat{\Theta}_n = (\widehat{Z}_n, \widehat{S}_n) = p_n(Z_n, S_n).$$

We will also denote Γ_n^Z the projection of Γ_n^Θ on E and Γ_n^S the projection of Γ_n^Θ on \mathbb{R}_+ . We use yet again new quantized approximations of operators H, I, K, J and L^d .

Definition 5.21. For $v \in L^2$, w defined on Γ_{n+1}^Z , $x \in E$, $n \in \{0, \dots, N-1\}$ and $z \in \Gamma_n^Z$, consider

$$\begin{aligned} \widehat{H}_{n+1}v(z, t) &= \mathbb{E}\left[e^{-\alpha(t \wedge t^*(\widehat{Z}_n))} v(\phi(\widehat{Z}_n, t \wedge t^*(\widehat{Z}_n))) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1} \geq t \wedge t^*(\widehat{Z}_n)\}} \Big| \widehat{Z}_n = z \right], \\ \widehat{I}_{n+1}w(z, t) &= \mathbb{E}\left[e^{-\alpha \widehat{S}_{n+1}} w(\widehat{Z}_{n+1}) \mathbf{1}_{\{\widehat{S}_{n+1} < t \wedge t^*(\widehat{Z}_n)\}} \Big| \widehat{Z}_n = z \right], \\ \widehat{K}_{n+1}v(z) &= \mathbb{E}\left[F(\widehat{Z}_n, t^*(\widehat{Z}_n)) + e^{-\alpha \widehat{S}_{n+1}} v(\widehat{Z}_{n+1}) \Big| \widehat{Z}_n = z \right], \\ \widehat{J}_{n+1}(v, w)(z, t) &= \mathbb{E}\left[F(\widehat{Z}_n, t) + e^{-\alpha \widehat{S}_{n+1}} w(\widehat{Z}_{n+1}) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1} < t \wedge t^*(\widehat{Z}_n)\}} \Big| \widehat{Z}_n = z \right] \\ &\quad + \mathbb{E}\left[e^{-\alpha(t \wedge t^*(\widehat{Z}_n))} v(\phi(\widehat{Z}_n, t \wedge t^*(\widehat{Z}_n))) \right. \\ &\quad \left. \times \mathbf{1}_{\{\widehat{S}_{n+1} \geq t \wedge t^*(\widehat{Z}_n)\}} \Big| \widehat{Z}_n = z \right], \\ \widehat{I}_{n+1}^d(v, w)(z) &= \min_{t \in G(z)} \widehat{J}_{n+1}(v, w)(z, t) \wedge \widehat{K}_{n+1}w(z). \end{aligned}$$

With these discretized operators and the previous evaluation of the \widehat{v}_k , we propose the following approximation scheme.

Definition 5.22. Consider $\widehat{v}_N(z) = g(z)$ where $z \in \Gamma_N^Z$ and for $k \in \{1, \dots, N\}$

$$\widehat{v}_{k-1}(z) = \widehat{I}_k^d(M\widehat{v}_k, \widehat{v}_k)(z), \tag{12}$$

where $z \in \Gamma_{k-1}^Z$.

See Fig. 1 for a graphical illustration of this numerical procedure. Therefore $\widehat{v}_0(\widehat{Z}_0)$ will be an approximation of $v_0(Z_0) = v_0(x_0)$. The derivation of the error bound for this scheme follows exactly the same lines as in the preceding section. Therefore we omit it and only state our main result.

Theorem 5.23. For all $0 \leq n \leq N-1$, suppose that $\Delta(z)$ for $z \in \Gamma_n^Z$ is such that

$$\sqrt{\frac{D^4 \|Z_n - \widehat{Z}_n\|_p + D^5 \|S_{n+1} - \widehat{S}_{n+1}\|_p}{D^3}} < \min_{z \in \Gamma_n^Z} \{\Delta(z)\}.$$

Then we have

$$\begin{aligned} &\|v_n(Z_n) - \widehat{v}_n(\widehat{Z}_n)\|_p \\ &\leq \|v_{n+1}(Z_{n+1}) - \widehat{v}_{n+1}(\widehat{Z}_{n+1})\|_p + \max_{y \in \mathbb{U}} |v_{n+1}(y) - \widehat{v}_{n+1}(y)| \\ &\quad + D_n^1 \|Z_n - \widehat{Z}_n\|_p + 3[v_{n+1}] \|Z_{n+1} - \widehat{Z}_{n+1}\|_p \\ &\quad + 2C_f \|S_{n+1} - \widehat{S}_{n+1}\|_p + D_n^2 \|\Delta(\widehat{Z}_n)\|_p \\ &\quad + 2\sqrt{D^3 \left(D^4 \|Z_n - \widehat{Z}_n\|_p + D^5 \|S_{n+1} - \widehat{S}_{n+1}\|_p \right)}, \end{aligned}$$

with

$$\begin{aligned} D_n^1 &= [v_n] + [Q][v_{n+1}]_1 \frac{C_\lambda}{\alpha} + \frac{C_f}{\alpha} (E_1 + E_2) \\ &\quad + \left\{ [Q][v_{n+1}]_* + 2E_3 \right\} \vee \left\{ 2([c]_1 + [c]_2[t^*]) + C_c E_1 \right. \\ &\quad \left. + \alpha[t^*] \left(\frac{C_f}{\alpha} + C_c \right) \right\}, \\ D_n^2 &= C_f + C_{v_{n+1}} C_\lambda + [c]_2 + (C_c + C_{v_{n+1}})(C_\lambda + \alpha), \end{aligned}$$

$$D^3 = \left(\frac{2C_f}{\alpha} + C_c \right) C_\lambda,$$

$$D^4 = 2[t^*] \left(\frac{2C_f}{\alpha} + C_c \right),$$

$$D^5 = 2 \left(\frac{2C_f}{\alpha} + C_c \right).$$

Remark 5.24. By using the same arguments as in Remark 5.12, it can be shown that $\|v_n(Z_n) - \widehat{v}_n(\widehat{Z}_n)\|_p$ can be made arbitrarily small by an adequate choice of the discretization parameters. From a theoretical point of view, the error can be calculated by iterating the result of Theorem 5.23. However, this result is not presented here because it would lead to an intricate expression. From a numerical point of view, a computer can easily estimate this error as shown in the example of Section 6.

5.4. Step by step description of the algorithm

Recall that the main objective of our algorithm is to compute the approximation $\widehat{v}_0(x_0)$ of the value function of the impulse control problem $v_0(x_0)$. The global recursive procedure is described on Fig. 1.

The calculation of $\widehat{v}_0(x_0)$ is based on the backward recursion given in Definition 5.22 and described in the first line of Fig. 1. It involves the operators \widehat{I}_j^d constructed with the quantized process $\widehat{\Theta}_n$ starting from x_0 . This recursion is not self contained and requires previous evaluation of the functions \widehat{v}_j on the control set \mathbb{U} .

The lower part of Fig. 1, shows how to compute these functions \widehat{v}_j at each point of the control grid \mathbb{U} . This is the triangular backward recursion given in Definition 5.9. More precisely, define $\widehat{v}_N = g$ and set $j < N$ and suppose that all the \widehat{v}_l for all $l > j$ have already been computed everywhere on the control set \mathbb{U} . One then computes \widehat{v}_j in the following way, following the j -th line of Fig. 1 counting from the bottom. One first iterates the operators $\widehat{I}_k^{1,d}$ and uses the quantized process $\widehat{\Theta}_n^1$, to obtain $\widehat{v}_j(y^1)$. Then one iterates the operators $\widehat{I}_k^{2,d}$ and uses the quantized process $\widehat{\Theta}_n^2$, to obtain $\widehat{v}_j(y^2)$, and so on until the last point $\widehat{v}_j(y^u)$. Thus one obtains \widehat{v}_j at all points of the control set \mathbb{U} .

5.5. Practical implementation

The procedure defined above is the natural one to obtain convergence rates for our approximations. However, in practice we proceed in a different order.

The first step is to fix the computational horizon N . This point was discussed earlier. The second step is not the time discretization, but the computation of the quantized approximations of the sequences (Θ_n) and (Θ_n^i) . The quantization algorithm may be quite long to run. However, it must be pointed out that this quantization step only depends on the optimization procedure through the control set \mathbb{U} but it does not depend on the cost functions f and c . The sequence $(\widehat{\Theta}_n)$ is obtained in a straightforward way. As for the $(\widehat{\Theta}_n^i)$, if the control set is very small, it is possible to run as many sequences of grids as there are points in the control set. Otherwise, one can do with only one sequence of grids computed with the Markov chain (Θ_n^μ) with a random starting point $Z_0 = Z_0^\mu$ uniformly distributed on the control set \mathbb{U} . To derive the point-wise approximation error, one simply uses the finiteness of \mathbb{U} and the definition of the L_p norm.

$$\begin{aligned} |v_k(y^j) - \widehat{v}_k(y^j)| &\leq u \sum_{i=1}^u |v_k(y^i) - \widehat{v}_k(y^i)| \frac{1}{u} \\ &\leq u^{1/p} \|v_k(Z_0^\mu) - \widehat{v}_k(\widehat{Z}_0^\mu)\|_p \end{aligned}$$

where u is the cardinal of \mathbb{U} . Notice that the last term is bounded in Theorem 5.11. Hence, one really only needs two series of quantization grids.

Once the quantization grids are computed and stored, one computes the path-adapted time grids $G(z)$ for all z in all the quantization grids, that is only a finite number of z . The step $\Delta(z)$ can usually be chosen constant equal to Δ , so that either one can store the whole time grids, or one only needs to store the values of Δ and $t^*(z)$ for all z in the quantization grids.

Once these preliminary computations are done, one can finally compute the value function. This last step is comparatively faster. The only point left to discussion is how to choose the initializing function g . The most interesting starting function is the cost h of the no impulse strategy, because then the value function \mathcal{V}_N has a natural interpretation. However, in general, one needs additional assumptions on Q to ensure that h is in \mathbf{L} . Another problem, is that in general computing h is a difficult problem, especially as we need to know its value at many different points, as explained in Remark 5.1. To overcome these difficulties, one can choose g to be an upper bound of h , for instance, $g = \alpha^{-1}C_f$. In the special cases where h can be explicitly computed, we advise to use h .

6. Example

Now we apply our procedure to a simple PDMP and present numerical results. This example is quite similar to example (54.29) in Davis (1993), we only added random jumps to obtain a non trivial Markov chain (Z_n, S_n) .

Set $E = [0, 1]$, and $\partial E = \{1\}$. The flow is defined on $[0, 1]$ by $\phi(x, t) = x + vt$ for some positive v , the jump rate is defined on $[0, 1]$ by $\lambda(x) = \beta x$, with $\beta > 0$, and for all $x \in [0, 1]$, one sets $Q(x, \cdot)$ to be the uniform law on $[0, 1/2]$. Thus, the process moves with constant speed v towards 1, but the closer it gets to the boundary 1, the higher the probability to jump backwards on $[0, 1/2]$. Figs. 2 and 3 show two trajectories of this process for $x_0 = 0$, $v = 1$ and $\beta = 3$ and up to the 10-th jump.

The running cost is defined on E by $f(x) = 1 - x$ and the intervention cost is a constant c_0 . Therefore, the best performance is obtained when the process is close to the boundary 1. The control set \mathbb{U} is the set of $\frac{k}{u}$, $0 \leq k \leq u - 1$ for some fixed integer u . In this special case, the control grid is already a discretization of the whole state space of the process. Therefore one needs only one series of grids starting from the control points to obtain an approximation of the value function at each point of the control grid.

We ran our algorithm for the parameters $x_0 = 0$, $v = 1$, $\beta = 3$, $c_0 = 0.08$, the discount factor $\alpha = 2$ and $u = 50$ points in the control grid and several values of the horizon N .

For an horizon $N = 5$ (respectively, $N = 10$, $N = 15$) interventions or jumps, Fig. 4 (respectively, Figs. 5 and 6) gives the approximated value function we obtained (computed at the 50 points of the control grid) for 50, 100 and 500 discretization points in each quantization grid and. As expected, the approximation gets smoother and lower as the number of points in the quantization grids increases.

The theoretical errors corresponding to the horizon $N = 5$ (respectively, $N = 10$, $N = 15$) are given in Table 1 (respectively, Tables 2 and 3). The values of the error are fairly high and conservative, but it must be pointed out that on the one hand, they do decrease as the number of points in the quantization grids increase, as expected; on the other hand their expressions are calculated and valid for a very wide and general class of PDMPs, hence when applied to a specific example, they cannot be very sharp.

Notice also that the approximated value function obtained for the horizon of $N = 10$ jumps or interventions is much lower than that obtained for the horizon $N = 5$ jumps or interventions.

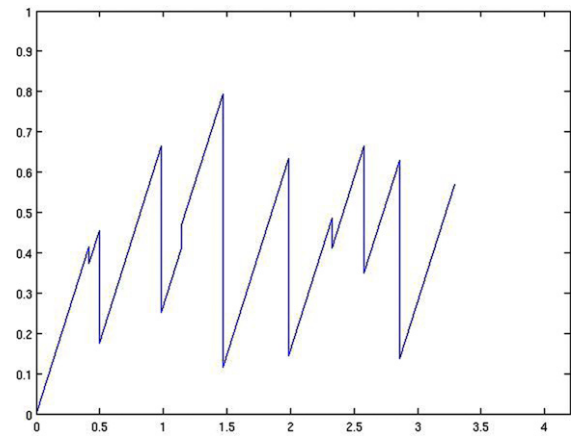


Fig. 2. One trajectory of the PDMP.

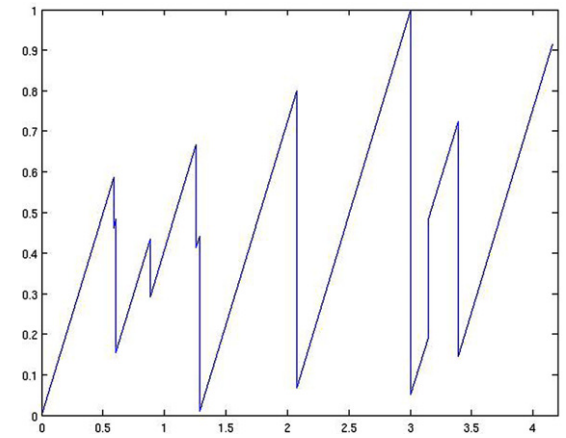


Fig. 3. Another trajectory of the PDMP.

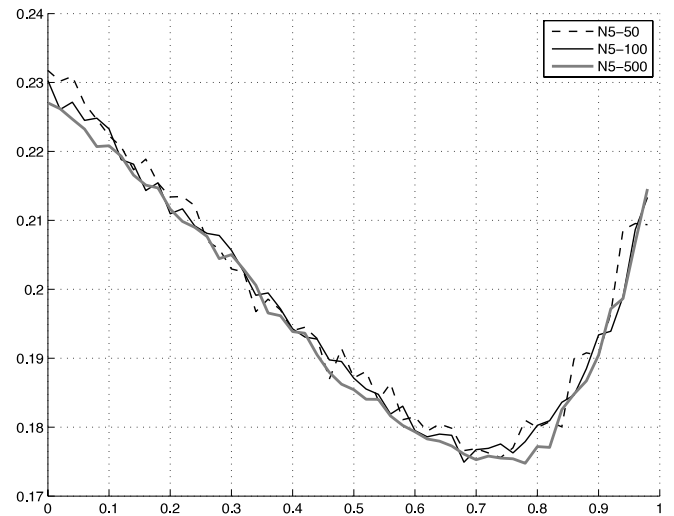


Fig. 4. Approximated value function for $N = 5$.

Table 1
Theoretical errors for $N = 5$.

Number of points in the quant. grids	$\ v_0(Z_0) - \hat{v}_0(Z_0)\ _2$
50	4636
100	3700
500	2141

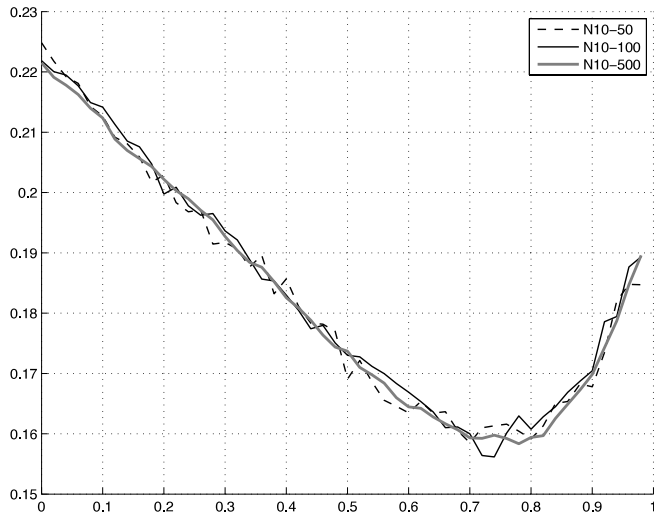


Fig. 5. Approximated value function for $N = 10$.

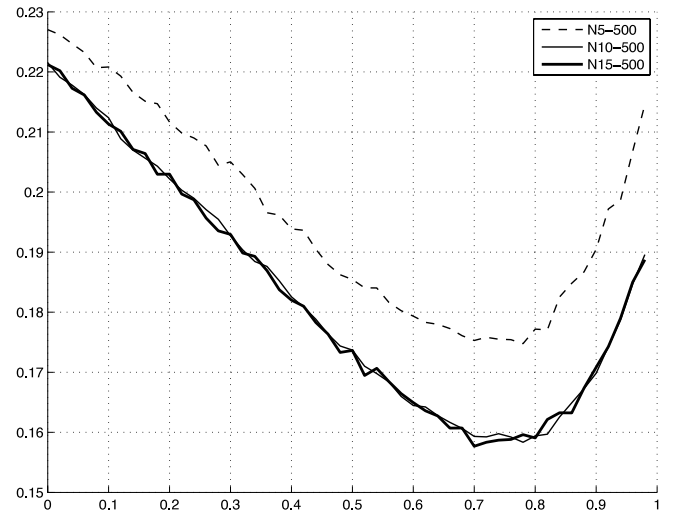


Fig. 7. Approximated value function for $N = 5, N = 10$ and $N = 15$ for 500 points in the quantization grids.

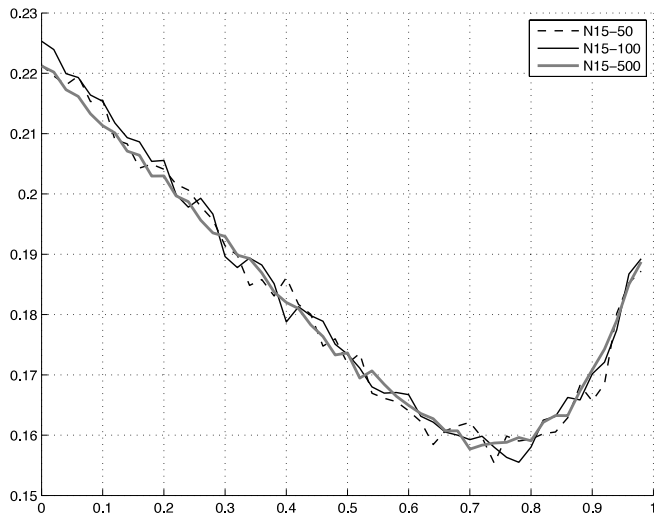


Fig. 6. Approximated value function for $N = 15$.

Table 2
Theoretical errors for $N = 10$.

Number of points in the quant. grids	$\ v_0(Z_0) - \widehat{v}_0(Z_0)\ _2$
50	$5.341 \cdot 10^{11}$
100	$4.501 \cdot 10^{11}$
500	$2.567 \cdot 10^{11}$

Table 3
Theoretical errors for $N = 15$.

Number of points in the quant. grids	$\ v_0(Z_0) - \widehat{v}_0(Z_0)\ _2$
50	$1.460 \cdot 10^{21}$
100	$1.288 \cdot 10^{21}$
500	$0.750 \cdot 10^{21}$

This is natural as it is a minimization problem, and the more there are possible interventions the lower the value function is. This also suggests that the horizon $N = 5$ is not large enough

to approximate the infinite horizon problem. Fig. 7 gives the approximated value function we obtained (computed at the 50 points of the control grid) for 500 points in the quantization grids and respective horizons of $N = 5, N = 10$ and $N = 15$ jumps or interventions. There is very little difference between the results for $N = 10$ and $N = 15$, suggesting that it is enough to take an horizon of 10 jumps or intervention to approximate the infinite horizon optimization problem.

Appendix. Lipschitz continuity results

A.1. Lipschitz properties of the operators

We start with preliminary results on operators M, H, F and I .

Lemma A.1. For any g defined on \mathbb{U} , $Mg \in \mathbf{L}$. Moreover,

$$[Mg]_1 \leq [c]_1, \quad [Mg]_2 \leq [c]_2, \quad [Mg]_* \leq [c]_*,$$

$$C_{Mg} \leq C_c + C_g.$$

Proof. By using the fact that $|Mg(x) - Mg(y)| \leq \sup_{z \in \mathbb{U}} |c(x, z) - c(y, z)|$ and Assumption 4.2, the result follows easily. \square

Lemma A.2. Let $v \in \mathbf{L}$. Then for all $(x, y) \in E^2$ and $(t, u) \in \mathbb{R}_+^2$, one has

$$|Hv(x, t) - Hv(y, u)| \leq D_1(v) |x - y| + D_2(v) |t - u|,$$

where

- if $t < t^*(x)$ and $u < t^*(y)$,

$$D_1(v) = [v]_1 + C_v C_{t^*} [\lambda]_1, \quad D_2(v) = [v]_2 + C_v (C_\lambda + \alpha),$$
- if $t = t^*(x)$ and $u = t^*(y)$,

$$D_1(v) = [v]_* + C_v (C_{t^*} [\lambda]_1 + (C_\lambda + \alpha) [t^*]),$$

$$D_2(v) = 0,$$
- otherwise,

$$D_1(v) = [v]_1 + [v]_2 [t^*] + C_v (C_{t^*} [\lambda]_1 + (C_\lambda + \alpha) [t^*]),$$

$$D_2(v) = [v]_2 + C_v (C_\lambda + \alpha).$$

Proof. This is straightforward. \square

Lemma A.3. For all $w \in \mathbf{L}$, $(x, y) \in E^2$ and $(t, u) \in \mathbb{R}_+^2$, one has

$$\begin{aligned} |F(x, t) - F(y, u)| &\leq \frac{1}{\alpha} \left([f]_1 + C_f C_{t^*} [\lambda]_1 \right) |x - y| \\ &\quad + C_f (|t - u| \vee [t^*] |x - y|), \\ |Iw(x, t) - Iw(y, u)| &\leq \frac{1}{\alpha} \left([Q][w]_1 C_\lambda \right. \\ &\quad \left. + C_w [\lambda]_1 (1 + C_\lambda C_{t^*}) \right) |x - y| \\ &\quad + C_w C_\lambda (|t - u| \vee [t^*] |x - y|). \end{aligned}$$

Proof. Suppose, without loss of generality, that $t \wedge t^*(x) \leq u \wedge t^*(y)$. Then, one has

$$\begin{aligned} |F(x, t) - F(y, u)| &\leq \int_0^{t \wedge t^*(x)} e^{-\alpha s} |f(\phi(x, s)) e^{-\Lambda(x, s)} \\ &\quad - f(\phi(y, s)) e^{-\Lambda(y, s)}| ds \\ &\quad + \int_{t \wedge t^*(x)}^{u \wedge t^*(y)} |f(\phi(y, s)) e^{-\alpha s - \Lambda(y, s)}| ds \\ &\leq \left([f]_1 + C_f C_{t^*} [\lambda]_1 \right) \int_0^\infty e^{-\alpha s} ds |x - y| \\ &\quad + C_f |u \wedge t^*(y) - t \wedge t^*(x)|. \end{aligned}$$

From the fact that $|u \wedge t^*(y) - t \wedge t^*(x)| \leq |t - u| \vee [t^*] |x - y|$ we get the first inequality. By using similar arguments, it is easy to obtain the last result. \square

Now we turn to the Lipschitz property of operator K .

Lemma A.4. For $w \in \mathbf{L}$ and $(x, y) \in E^2$, one has

$$\begin{aligned} |Kw(x) - Kw(y)| &\leq \left\{ [Q][w]_1 \frac{C_\lambda}{\alpha} + [Q][w]_* \right. \\ &\quad \left. + C_w (E_1 + E_2) + E_3 \right\} |x - y|. \end{aligned}$$

Proof. This is a direct consequence of the definition of operator K and Lemmas A.2 and A.3. \square

Finally, we study the Lipschitz properties of operator J .

Lemma A.5. For all $(v, w) \in \mathbf{C}^2$, $x \in E$ and $(t, u) \in \mathbb{R}_+^2$, one has

$$\begin{aligned} |J(v, w)(x, t) - J(v, w)(x, u)| &\leq \left(C_f + C_w C_\lambda \right. \\ &\quad \left. + [v]_2 + C_v (C_\lambda + \alpha) \right) |t - u|. \end{aligned}$$

Proof. By using (1) and Lemmas A.2 and A.3, the result follows easily. \square

Lemma A.6. For all $(v, w) \in \mathbf{L}^2$, $(x, y) \in E^2$ and $t \geq 0$, one has

$$\begin{aligned} |J(v, w)(x, t) - J(v, w)(y, t)| &\leq \left\{ [v]_1 + [v]_2 [t^*] \right. \\ &\quad \left. + [Q][w]_1 \frac{C_\lambda}{\alpha} + C_v E_1 + C_w E_2 + E_3 \right\} |x - y| \end{aligned}$$

where

$$\begin{aligned} E_1 &= C_{t^*} [\lambda]_1 + (C_\lambda + \alpha) [t^*], \\ E_2 &= C_\lambda [t^*] + [\lambda]_1 \frac{1 + C_\lambda C_{t^*}}{\alpha}, \\ E_3 &= [f]_1 \frac{1}{\alpha} + C_f \left(\frac{C_{t^*} [\lambda]_1}{\alpha} + [t^*] \right). \end{aligned}$$

Proof. Again, this is a direct consequence of (1) and Lemmas A.2 and A.3. \square

Remark A.7. It is easy to obtain that for $(v, w) \in \mathbf{C}^2$, $s \in \mathbb{R}_+$ and $(x, y) \in E^2$,

$$\begin{aligned} \left| \inf_{t \geq s} J(v, w)(x, t) - \inf_{t \geq s} J(v, w)(y, t) \right| \\ \leq \sup_{t \geq 0} |J(v, w)(x, t) - J(v, w)(y, t)|. \end{aligned}$$

Lemma A.8. Let $(v, w) \in \mathbf{L}^2$. Then for all $x \in E$ and $(s, t) \in \mathbb{R}_+^2$,

$$\begin{aligned} \left| \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u) \right| \\ \leq \left(C_f + C_w C_\lambda + [v]_2 + C_v (C_\lambda + \alpha) \right) |t - s|. \end{aligned}$$

Proof. Without loss of generality it can be assumed that $s \leq t$. Therefore, one has

$$\begin{aligned} \left| \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u) \right| \\ = \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u). \end{aligned} \tag{A.1}$$

Remark that there exists $\bar{s} \in [s \wedge t^*(x), t^*(x)]$ such that $\inf_{u \geq s} J(v, w)(x, u) = J(v, w)(x, \bar{s})$. Consequently, if $\bar{s} \geq t \wedge t^*(x)$ then one has $\left| \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u) \right| = 0$.

Now if $\bar{s} \in [s \wedge t^*(x), t \wedge t^*(x)]$, then one has

$$\begin{aligned} \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u) \\ \leq J(v, w)(x, t) - J(v, w)(x, \bar{s}). \end{aligned}$$

From Lemma A.5, we obtain the following inequality

$$\begin{aligned} \inf_{u \geq t} J(v, w)(x, u) - \inf_{u \geq s} J(v, w)(x, u) \\ \leq \left(C_f + C_w C_\lambda + [v]_2 + C_v (C_\lambda + \alpha) \right) |t - \bar{s}|. \end{aligned} \tag{A.2}$$

Combining Eqs. (A.1) and (A.2) and the fact that $|t - \bar{s}| \leq |t - s|$ the result follows. \square

A.2. Lipschitz properties of the operator \mathcal{L}

Now we study the Lipschitz continuity of our main operator.

Lemma A.9. For all $(v, w) \in \mathbf{L}^2$, $x \in E$ and $t \in [0, t^*(x)]$ and $u \in \mathbb{R}_+$, one has

$$\begin{aligned} F(\phi(x, t), u) &= e^{\alpha t + \Lambda(x, t)} (F(x, t + u) - F(x, u)), \\ Iw(\phi(x, t), u) &= e^{\alpha t + \Lambda(x, t)} (Iw(x, t + u) - Iw(x, u)), \\ Hv(\phi(x, t), u) &= e^{\alpha t + \Lambda(x, t)} Hv(x, t + u). \end{aligned}$$

Proof. By using the semi-group property of ϕ , we have $\Lambda(\phi(x, t), u) = \Lambda(x, t + u) - \Lambda(x, t)$ for $t + u < t^*(x)$ and noting that

$t^*(\phi(x, t)) = t^*(x) - t$ for $t < t^*(x)$, a simple change of variable yields

$$F(\phi(x, t), u) = e^{\alpha t + \Lambda(x,t)} \int_t^{(t+u) \wedge t^*(x)} e^{-\alpha s - \Lambda(x,s)} f(\phi(x, s)) ds,$$

and we get the first equation. The other equalities can be obtained by using similar arguments. \square

Lemma A.10. For all $(v, w) \in \mathbf{L}^2, x \in E$ and $t \in [0, t^*(x)]$,

$$L(v, w)(\phi(x, t)) = e^{\alpha t + \Lambda(x,t)} \left[\left\{ \inf_{s \geq t} J(v, w)(x, s) \wedge Kw(x) \right\} - F(x, t) - Iw(x, t) \right].$$

Proof. For $t \in [0, t^*(x)]$ and $u \in \mathbb{R}_+$, we have from Lemma A.9, (1)

$$J(v, w)(\phi(x, t), u) = e^{\alpha t + \Lambda(x,t)} [J(v, w)(x, t+u) - F(x, t) - Iw(x, t)],$$

$$Kw(\phi(x, t)) = e^{\alpha t + \Lambda(x,t)} [Kw(x) - F(x, t) - Iw(x, t)].$$

Consequently, from Eq. (2), it follows

$$L(v, w)(\phi(x, t)) = e^{\alpha t + \Lambda(x,t)} \left[\left\{ \inf_{u \geq 0} J(v, w)(x, t+u) \wedge Kw(x) \right\} - F(x, t) - Iw(x, t) \right],$$

showing the result. \square

Proposition A.11. For all $w \in \mathbf{L}, \mathcal{L}w \in \mathbf{L}$, and

$$\begin{aligned} [\mathcal{L}w]_1 &\leq e^{(\alpha+C_\lambda)C_t^*} \left\{ \left(\frac{C_f}{\alpha} \right) + ([c]_1 + [c]_2[t^*] + C_c E_1) \vee ([Q][w]_*) \right. \\ &\quad \left. + [\lambda]_1 C_t^* C_c + 2E_3 + \frac{2[Q]C_\lambda}{\alpha} [w]_1 \right. \\ &\quad \left. + \{E_1 + 2E_2 + [\lambda]_1 C_t^* (1 + C_\lambda/\alpha)\} C_w \right\}, \\ [\mathcal{L}w]_2 &\leq e^{(\alpha+C_\lambda)C_t^*} \left\{ 3C_f + [c]_2 + 2C_c(C_\lambda + \alpha) + \frac{C_f C_\lambda}{\alpha} \right. \\ &\quad \left. + C_w \left[4C_\lambda + \frac{C_\lambda^2}{\alpha} + \alpha \right] \right\}, \\ [\mathcal{L}w]_* &\leq [\mathcal{L}w]_1 + [\mathcal{L}w]_2 [t]_*, \\ [\mathcal{L}w] &\leq \frac{[Q]C_\lambda}{\alpha} [w]_1 + E_3 + ([c]_1 + [c]_2[t^*] + C_c E_1) \vee ([Q][w]_*) + \{E_1 + E_2\} C_w. \end{aligned}$$

Proof. Let us denote $\mathcal{L}w$ by g . We have for $(x, y) \in E^2$ and $t \in [0, t^*(x) \wedge t^*(y)]$

$$\begin{aligned} |g(\phi(x, t)) - g(\phi(y, t))| &\leq e^{\alpha t + \Lambda(y,t)} \left\{ |F(x, t) - F(y, t)| \right. \\ &\quad \left. + |Iw(x, t) - Iw(y, t)| \right\} \\ &\quad + e^{\alpha t + \Lambda(y,t)} \left\{ \left| \inf_{s \geq t} J(Mw, w)(x, s) \right. \right. \\ &\quad \left. \left. - \inf_{s \geq t} J(Mw, w)(y, s) \right| \right\} \end{aligned}$$

$$\begin{aligned} &\vee |Kw(x) - Kw(y)| \Big\} \\ &+ \left| e^{\alpha t + \Lambda(x,t)} - e^{\alpha t + \Lambda(y,t)} \right| \\ &\times \left\{ \left| \inf_{s \geq t} J(Mw, w)(x, s) \right. \right. \\ &\quad \left. \left. \wedge Kw(x) \right\} - F(x, t) - Iw(x, t) \right|. \end{aligned}$$

It is easy to show that for $x \in E, t \in [0, t^*(x)]$, and $w \in \mathbf{L}$ we have $e^{\alpha t + \Lambda(x,t)} \leq e^{(\alpha+C_\lambda)C_t^*}, \left\{ \inf_{s \geq t} J(Mw, w)(x, s) \wedge Kw(x) \right\} - F(x, t) - Iw(x, t) \leq \frac{1}{\alpha}(C_f + C_\lambda C_w) + C_c + C_w$ and for $(x, y) \in E^2$ and $t \in [0, t^*(x) \wedge t^*(y)] \left| e^{\alpha t + \Lambda(x,t)} - e^{\alpha t + \Lambda(y,t)} \right| \leq e^{(\alpha+C_\lambda)C_t^*} [\lambda]_1 C_t^* |x - y|$. Consequently, by using Lemmas A.3 and A.4 and Remark A.7, we get the first equation.

For $x \in E$ and $(s, t) \in [0, t^*(x)]^2$

$$\begin{aligned} &\left| g(\phi(x, s)) - g(\phi(x, t)) \right| \\ &\leq e^{\alpha t + \Lambda(x,t)} |Iw(x, s) - Iw(x, t)| \\ &\quad + e^{\alpha t + \Lambda(x,t)} \left\{ \left| \inf_{u \geq s} J(Mw, w)(x, u) - \inf_{u \geq t} J(Mw, w)(x, u) \right| \right. \\ &\quad \left. + |F(x, s) - F(x, t)| \right\} + \left| e^{\alpha s + \Lambda(x,s)} - e^{\alpha t + \Lambda(x,t)} \right| \\ &\quad \times \left\{ \inf_{u \geq s} J(Mw, w)(x, u) \wedge Kw(x) \right\} - F(x, s) - Iw(x, s). \end{aligned}$$

Note that for $x \in E, (s, t) \in [0, t^*(x)]^2 \left| e^{\alpha s + \Lambda(x,s)} - e^{\alpha t + \Lambda(x,t)} \right| \leq e^{(\alpha+C_\lambda)C_t^*} (C_\lambda + \alpha) |t - s|$. Consequently, by using Lemmas A.3 and A.8, we obtain the second equation.

The third inequality is straightforward and finally, for $(x, y) \in E^2$ we have

$$\begin{aligned} |g(x) - g(y)| &\leq \left| \inf_{s \geq 0} J(Mw, w)(x, s) \right. \\ &\quad \left. - \inf_{s \geq 0} J(Mw, w)(y, s) \right| \vee |Kw(x) - Kw(y)|. \end{aligned}$$

By using Remark A.7 and Lemma A.4, we get the last equation. \square

Corollary A.12. For all $0 \leq n \leq N$, the value functions v_n are in \mathbf{L} .

Proof. As $v_N = g$ is in \mathbf{L} by assumption, a recursive application of Proposition A.11 yields the result. \square

References

Bally, V., & Pagès, G. (2003). A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli*, 9(6), 1003–1049.

Bally, V., Pagès, G., & Printems, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Mathematical Finance*, 15(1), 119–168.

Costa, O. L. V. (1993). Discretizations for the average impulse control of piecewise deterministic processes. *Journal of Applied Probability*, 30(2), 405–420.

Costa, O. L. V., & Davis, M. H. A. (1988). Approximations for optimal stopping of a piecewise-deterministic process. *Mathematics of Control Signals and Systems*, 1(2), 123–146.

Costa, O. L. V., & Davis, M. H. A. (1989). Impulse control of piecewise-deterministic processes. *Mathematics of Control Signals and Systems*, 2(3), 187–206.

Davis, M. H. A. (1993). *Monographs on statistics and applied probability: Vol. 49. Markov models and optimization*. London: Chapman & Hall.

de Saporta, B., Dufour, F., & Gonzalez, K. (2010). Numerical method for optimal stopping of piecewise deterministic Markov processes. *Annals of Applied Probability*, 20(5), 1607–1637.

- Dempster, M. A. H., & Ye, J. J. (1995). Impulse control of piecewise deterministic Markov processes. *Annals of Applied Probability*, 5(2), 399–423.
- Gat̆arek, D. (1991). On first-order quasi-variational inequalities with integral terms. *Applied Mathematics and Optimization*, 24(1), 85–98.
- Gat̆arek, D. (1992). Optimality conditions for impulsive control of piecewise-deterministic processes. *Mathematics of Control Signals and Systems*, 5(2), 217–232.
- Gray, R. M., & Neuhoff, D. L. (1948–1998). Quantization. *IEEE Transactions on Information Theory*, 44(6), 2325–2383. Information theory.
- Kushner, H. J. (1977). *Mathematics in science and engineering: Vol. 129. Probability methods for approximations in stochastic control and for elliptic equations*. New York: Academic Press [Harcourt Brace Jovanovich Publishers].
- Lenhart, S. M. (1989). Viscosity solutions associated with impulse control problems for piecewise-deterministic processes. *International Journal of Mathematics and Mathematical Sciences*, 12(1), 145–157.
- Pagès, G. (1998). A space quantization method for numerical integration. *Journal of Computational and Applied Mathematics*, 89(1), 1–38.
- Pagès, G., & Pham, H. (2005). Optimal quantization methods for nonlinear filtering with discrete-time observations. *Bernoulli*, 11(5), 893–932.
- Pagès, G., Pham, H., & Printems, J. (2004a). An optimal Markovian quantization algorithm for multi-dimensional stochastic control problems. *Stochastics and Dynamics*, 4(4), 501–545.
- Pagès, G., Pham, H., & Printems, J. (2004b). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of computational and numerical methods in finance* (pp. 253–297). Boston, MA: Birkhäuser Boston.



Benoîte de Saporta received her Ph.D. degree in applied probability in 2004 from Rennes 1 University (France). Since 2006, she has been Assistant Professor at Montesquieu Bordeaux IV University, Department of Economics. She is also a member of the INRIA Team CQFD. Her research is focused on applied probability and more specifically estimation, control and numerical approximations for autoregressive processes, Markov switching processes and hybrid systems, with applications in reliability and biology.



François Dufour is Professor at the Institut Polytechnique of Bordeaux (France) since 2002. He is also the leader of INRIA team CQFD (quality control and dynamic reliability). He is associate editor of *SIAM Journal of Control and Optimization* since 2009. His research interests include Markov processes, stochastic stability and stochastic optimal control with applications in reliability and target tracking.

3 Articles relatifs au Chapitre 3

- 3.1 [BdSD12a] A. BRANDEJSKY, B. DE SAPORTA, AND F. DUFOUR. Numerical method for expectations of piecewise-deterministic markov processes. *CAMCoS*, 7(1) :63–104, 2012.

*Communications in
Applied
Mathematics and
Computational
Science*

**NUMERICAL METHOD FOR EXPECTATIONS OF
PIECEWISE DETERMINISTIC MARKOV
PROCESSES**

ADRIEN BRANDEJSKY, BENOÎTE DE SAPORTA
AND FRANÇOIS DUFOUR

vol. 7 no. 1 2012

NUMERICAL METHOD FOR EXPECTATIONS OF PIECEWISE DETERMINISTIC MARKOV PROCESSES

ADRIEN BRANDEJSKY, BENOÎTE DE SAPORTA AND FRANÇOIS DUFOUR

We present a numerical method to compute expectations of functionals of a piecewise deterministic Markov process. We discuss time dependent functionals as well as deterministic time horizon problems. Our approach is based on the quantization of an underlying discrete-time Markov chain. We obtain bounds for the rate of convergence of the algorithm. The approximation we propose is easily computable and is flexible with respect to some of the parameters defining the problem. An example illustrates the paper.

1. Introduction	63
2. Definitions and assumptions	67
3. Expectation	71
4. Approximation scheme	74
5. Time-dependent functionals	78
6. Numerical results	89
7. Conclusion	97
Appendix A. Lipschitz continuity of F , G and v_n	97
Appendix B. Relaxed assumption on the running cost function	101
Appendix C. Proof of Theorem 4.5	103
Acknowledgements	103
References	104

1. Introduction

The aim of this paper is to propose a practical numerical method to approximate some expectations related to a piecewise deterministic Markov process thanks to the quantization of a discrete-time Markov chain naturally embedded within the continuous-time process.

Piecewise deterministic Markov processes (PDMP's) have been introduced by M. H. A. Davis in [5] as a general class of stochastic models. PDMP's are a

This work was supported by ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004.

MSC2010: primary 60J25, 65C20; secondary 60K10.

Keywords: expectation, piecewise deterministic Markov processes, quantization, numerical method.

family of Markov processes involving deterministic motion punctuated by random jumps. The motion depends on three local characteristics namely the flow Φ , the jump rate λ and the transition measure Q , which specifies the postjump location. Starting from the point x , the motion of the process follows the flow $\Phi(x, t)$ until the first jump time T_1 , which occurs either spontaneously in a Poisson-like fashion with rate $\lambda(\Phi(x, t))$ or when the flow $\Phi(x, t)$ hits the boundary of the state space. In either case, the location of the process at the jump time T_1 is selected by the transition measure $Q(\Phi(x, T_1), \cdot)$ and the motion restarts from this new point X_{T_1} denoted by Z_1 . We define similarly the time S_2 until the next jump, $T_2 = T_1 + S_2$ with the next postjump location defined by $Z_2 = X_{T_2}$ and so on. Thus, associated to the PDMP we have the discrete-time Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$, given by the postjump locations and the interjump times. A suitable choice of the state space and the local characteristics Φ , λ and Q provides stochastic models covering a great number of problems of operations research as described in [5, Section 33].

We are interested in the approximation of expectations of the form

$$E_x \left[\int_0^{T_N} l(X_t) dt + \sum_{j=1}^N c(X_{T_j^-}) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right]$$

where $(X_t)_{t \geq 0}$ is a PDMP and l and c are some nonnegative, real-valued, bounded functions and ∂E is the boundary of the domain. Such expectations are discussed by M. H. A. Davis in [5, Chapter 3]. They often appear as cost or reward functions in optimization problems. The first term is referred to as the running cost while the second may be called the boundary jump cost. Besides, they are quite general since Davis shows how a “wide variety of apparently different functionals” can be obtained from the above specific form. For example, this wide variety includes quantities such as a mean exit time and even, for any fixed $t \geq 0$, the distribution of X_t (that is, $E_x[\mathbb{1}_F(X_t)]$ where F is a measurable set).

There are surprisingly few works in the literature devoted to the actual computation of such expectations, using other means than direct Monte Carlo simulations. Davis showed that these expectations satisfy integrodifferential equations. However, the set of partial differential equations that is obtained is unusual. Roughly speaking, these differential equations are basically transport equations with a non-constant velocity and they are coupled by the boundary conditions and by some integral terms involving kernels that are derived from the properties of the underlying stochastic process. The main difficulty comes from the fact that the domains on which the equations have to be solved vary from one equation to another making their numerical resolution highly problem specific. Another similar approach has been recently investigated in [4; 7]. It is based on a discretization of the Chapman Kolmogorov equations satisfied by the distribution of the process $(X_t)_{t \geq 0}$. The

authors propose an approximation of such expectations based on finite volume methods. Unfortunately, their method is only valid if there are no jumps at the boundary. Our approach is completely different and does not rely on differential equations, but on the fact that such expectations can be computed by iterating an integral operator G . This operator only involves the embedded Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$ and conditional expectations. It is therefore natural to propose a computational method based on the quantization of this Markov chain, following the same idea as [6].

There exists an extensive literature on quantization methods for random variables and processes. The interested reader may for instance consult [8], [9] and the references within. Quantization methods have been developed recently in numerical probability or optimal stochastic control with applications in finance (see [1; 2; 9], for instance). The quantization of a random variable X consists in finding a finite grid such that the projection \hat{X} of X on this grid minimizes some L^p norm of the difference $X - \hat{X}$. Roughly speaking, such a grid will have more points in the areas of high density of X . As explained for instance in [9, Section 3], under some Lipschitz-continuity conditions, bounds for the rate of convergence of functionals of the quantized process towards the original process are available.

In the present work, we develop a numerical method to compute expectations of functionals of the above form where the cost functions l and c satisfy some Lipschitz-continuity conditions. We first recall the results presented by Davis according to whom, the above expectation may be computed by iterating an operator denoted by G . Consequently, it appears natural to follow the idea developed in [6] namely to express the operator G in terms of the underlying discrete-time Markov chain $(Z_n, S_n)_{n \in \mathbb{N}}$ and to replace it by its quantized approximation. Moreover, in order to prove the convergence of our algorithm, we replace the indicator function $\mathbb{1}_{\{X_{T_j^-} \in \partial E\}}$ contained within the functional by some Lipschitz continuous approximation. Bounds for the rate of convergence are then obtained. However, and this is the main contribution of this paper, we then tackle two important aspects that had not been investigated in [6].

The first aspect consists in allowing c and l to be time-dependent functions, although still Lipschitz continuous, so that we may compute expectations of the form

$$E_x \left[\int_0^{T_N} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right].$$

This important generalization has huge applicative consequences. For instance, it allows discounted cost or reward functions such as $l(x, t) = e^{-\delta t} l(x)$ and $c(x, t) = e^{-\delta t} c(x)$ where δ is some interest rate. To compute the above expectation, our strategy consists in considering, as suggested by Davis in [5], the time-augmented

process $\tilde{X}_t = (X_t, t)$. Therefore, a natural way to deal with the time-dependent problem is to apply our previous approximation scheme to the time-augmented process $(\tilde{X}_t)_{t \geq 0}$. However, it is far from obvious, that the assumptions required by our numerical method still hold for this new PDMP $(\tilde{X}_t)_{t \geq 0}$.

The second important generalization is to consider the deterministic time horizon problem. Indeed, it seems crucial, regarding the applications, to be able to approximate

$$\mathbf{E}_x \left[\int_0^{t_f} l(X_t, t) dt + \sum_{T_j \leq t_f} c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right]$$

for some fixed $t_f > 0$ regardless of how many jumps occur before this deterministic time. To compute this quantity, we start by choosing a time N such that $\mathbf{P}(T_N < t_f)$ be small so that the previous expectation boils down to

$$\mathbf{E}_x \left[\int_0^{T_N} l(X_t, t) \mathbb{1}_{\{t \leq t_f\}} dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} \right].$$

At first sight, this functional seems to be of the previous form. Yet, one must recall that Lipschitz continuity conditions have been made concerning the cost functions so that the indicator functions $\mathbb{1}_{\{t \leq t_f\}}$ prevent a direct application of the earlier results. We deal with the two indicator functions in two different ways. On the one hand, we prove that it is possible to relax the regularity condition on the running cost function so that our algorithm still converges in spite of the first indicator function. On the other hand, since the same reasoning cannot be applied to the indicator function within the boundary jump cost term, we bound it between two Lipschitz continuous functions. This provides bounds for the expectation of the deterministic time horizon functional.

An important advantage of our method is that it is flexible. Indeed, as pointed out in [1], a quantization based method is ‘‘obstacle free’’ which means, in our case, that it produces, once and for all, a discretization of the process independently of the functions l and c since the quantization grids merely depend on the dynamics of the process. They are only computed once, stored off-line and may therefore serve many purposes. Once they have been obtained, we are able to approximate very easily and quickly any of the expectations described earlier. This flexibility is definitely an important advantage of our scheme over standard methods such as Monte Carlo simulations since, with such methods, we would have to run the whole algorithm for each expectation we want to compute. This point is illustrated in Section 6 where we easily solve an optimization problem that would be very laboriously handled by Monte Carlo simulations.

The paper is organized as follows. We first recall, in Section 2, the definition of a PDMP and state our assumptions. In Section 3, we introduce the recursive method to compute the expectation. Section 4 presents the approximation scheme and a bound for the rate of convergence. The main contribution of the paper lies in Section 5, which contains generalizations to time-dependent parameters and deterministic time-horizon problems. The paper is illustrated by a numerical example in Section 6; a conclusion (Section 7) is followed by some appendixes containing technical results.

2. Definitions and assumptions

For all metric space E , we denote by $\mathcal{B}(E)$ its Borel σ -field and $B(E)$ the set of real-valued, bounded and measurable functions defined on E . For $a, b \in \mathbb{R}$, set $a \wedge b = \min(a, b)$, $a \vee b = \max(a, b)$, and $a^+ = a \vee 0$.

Definition of a PDMP. In this first section, let us define a piecewise deterministic Markov process and introduce some general assumptions. Let M be a finite set called the set of the modes that will represent the different regimes of evolution of the PDMP. For each $m \in M$, the process evolves in E_m , an open subset of \mathbb{R}^d . Let

$$E = \{(m, \zeta), m \in M, \zeta \in E_m\}.$$

This is the state space of the process $(X_t)_{t \in \mathbb{R}^+} = (m_t, \zeta_t)_{t \in \mathbb{R}^+}$. Let ∂E be its boundary and \bar{E} its closure and for any subset Y of E , Y^c denotes its complement.

A PDMP is defined by its local characteristics $(\Phi_m, \lambda_m, Q_m)_{m \in M}$.

- For each $m \in M$, $\Phi_m : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is a continuous function called the flow in mode m . For all $t \in \mathbb{R}$, $\Phi_m(\cdot, t)$ is an homeomorphism and $t \rightarrow \Phi_m(\cdot, t)$ is a semigroup; i.e., for all $\zeta \in \mathbb{R}^d$, $\Phi_m(\zeta, t + s) = \Phi_m(\Phi_m(\zeta, s), t)$. For all $x = (m, \zeta) \in E$, define the deterministic exit time from E :

$$t^*(x) = \inf \{t > 0 \text{ such that } \Phi_m(\zeta, t) \in \partial E_m\}.$$

We use here and throughout the convention $\inf \emptyset = +\infty$.

- For all $m \in M$, the jump rate $\lambda_m : \bar{E}_m \rightarrow \mathbb{R}^+$ is measurable, and for all $(m, \zeta) \in E$, there exists $\epsilon > 0$ such that

$$\int_0^\epsilon \lambda_m(\Phi_m(\zeta, t)) dt < +\infty.$$

- For all $m \in M$, Q_m is a Markov kernel on $(\mathcal{B}(\bar{E}), \bar{E}_m)$ that satisfies

$$Q_m(\zeta, \{(m, \zeta)\}^c) = 1 \quad \text{for all } \zeta \in \bar{E}_m.$$

From these characteristics, it can be shown (see [5]) that there exists a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, (\mathbf{P}_x)_{x \in E})$ on which a process $(X_t)_{t \in \mathbb{R}^+}$ is defined. Its motion, starting from a point $x \in E$, may be constructed as follows. Let T_1 be a nonnegative random variable with survival function

$$\mathbf{P}_x(T_1 > t) = \begin{cases} e^{-\Lambda(x,t)} & \text{if } 0 \leq t < t^*(x), \\ 0 & \text{if } t \geq t^*(x), \end{cases}$$

where for $x = (m, \zeta) \in E$ and $t \in [0, t^*(x)]$,

$$\Lambda(x, t) = \int_0^t \lambda_m(\Phi_m(\zeta, s)) ds.$$

One then chooses an E -valued random variable Z_1 according to the distribution $Q_m(\Phi_m(\zeta, T_1), \cdot)$. The trajectory of X_t for $t \leq T_1$ is

$$X_t = \begin{cases} (m, \Phi_m(\zeta, t)) & \text{if } t < T_1, \\ Z_1 & \text{if } t = T_1. \end{cases}$$

Starting from the point $X_{T_1} = Z_1$, one then selects in a similar way $S_2 = T_2 - T_1$ the time between T_1 and the next jump, Z_2 the next postjump location and so on. Davis shows, in [5], that the process so defined is a strong Markov process $(X_t)_{t \geq 0}$ with jump times $(T_n)_{n \in \mathbb{N}}$ (with $T_0 = 0$). The process $(\Theta_n)_{n \in \mathbb{N}} = (Z_n, S_n)_{n \in \mathbb{N}}$ where $Z_n = X_{T_n}$ is the postjump location and $S_n = T_n - T_{n-1}$ (with $S_0 = 0$) is the n -th interjump time is clearly a discrete-time Markov chain.

The following assumption about the jump-times is standard (see [5, Section 24], for example):

Assumption 2.1. For all $(x, t) \in E \times \mathbb{R}^+$, $\mathbf{E}_x[\sum_k \mathbb{1}_{\{T_k < t\}}] < +\infty$.

It implies in particular that T_k goes to infinity a.s. when k goes to infinity.

Notation and assumptions. For notational convenience, any function h defined on E will be identified with its component functions h_m defined on E_m . Thus, one may write

$$h(x) = h_m(\zeta) \text{ when } x = (m, \zeta) \in E.$$

We also define a generalized flow $\Phi : E \times \mathbb{R}^+ \rightarrow E$ such that

$$\Phi(x, t) = (m, \Phi_m(\zeta, t)) \text{ when } x = (m, \zeta) \in E.$$

Define on E the following distance, for $x = (m, \zeta)$ and $x' = (m', \zeta') \in E$:

$$|x - x'| = \begin{cases} +\infty & \text{if } m \neq m', \\ |\zeta - \zeta'| & \text{otherwise.} \end{cases} \quad (1)$$

For any function w in $B(\bar{E})$, introduce the notation

$$Qw(x) = \int_E w(y)Q(x, dy), \quad C_w = \sup_{x \in \bar{E}} |w(x)|,$$

and for any Lipschitz continuous function w in $B(E)$, denote by $[w]^E$, or if there is no ambiguity by $[w]$, its Lipschitz constant:

$$[w]^E = \sup_{x \neq y \in E} \frac{|w(x) - w(y)|}{|x - y|},$$

with the convention $\frac{1}{\infty} = 0$.

Remark 2.2. For $w \in B(\bar{E})$ and from the definition of the distance on E , one has $[w] = \max_{m \in M} [w_m]$.

Definition 2.3. Denote by $L_c(E)$ the set of functions $w \in B(E)$ that are Lipschitz continuous along the flow; i.e., the real-valued, bounded, measurable functions defined on E and satisfying the following conditions:

- For all $x \in E$, the map $w(\Phi(x, \cdot)) : [0, t^*(x)) \rightarrow \mathbb{R}$ is continuous, and the limit $\lim_{t \rightarrow t^*(x)} w(\Phi(x, t))$ exists and is denoted by $w(\Phi(x, t^*(x)))$.
- There exists $[w]_1^E \in \mathbb{R}^+$ such that for all $x, y \in E$ and $t \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|w(\Phi(x, t)) - w(\Phi(y, t))| \leq [w]_1^E |x - y|.$$

- There exists $[w]_2^E \in \mathbb{R}^+$ such that for all $x \in E$ and $t, u \in [0, t^*(x)]$, one has

$$|w(\Phi(x, t)) - w(\Phi(x, u))| \leq [w]_2^E |t - u|.$$

- There exists $[w]_*^E \in \mathbb{R}^+$ such that for all $x, y \in E$, one has

$$|w(\Phi(x, t^*(x))) - w(\Phi(y, t^*(y)))| \leq [w]_*^E |x - y|.$$

Denote by $L_c(\partial E)$ the set of real-valued, bounded, measurable functions defined on ∂E satisfying the following condition:

- There exists $[w]_*^{\partial E} \in \mathbb{R}^+$ such that for all $x, y \in E$, one has

$$|w(\Phi(x, t^*(x))) - w(\Phi(y, t^*(y)))| \leq [w]_*^{\partial E} |x - y|.$$

Remark 2.4. When there is no ambiguity, we will use the notation $[w]_i$ instead of $[w]_i^E$ for $i \in \{1, 2, *\}$ and $[w]_*$ instead of $[w]_*^E$.

Remark 2.5. In Definition 2.3, we used the generalized flow for notational convenience. For instance, the definition of $[w]_1$ is equivalent to the following: for all

$m \in M$, there exists $[w_m]_1 \in \mathbb{R}^+$ such that for all $\zeta, \zeta' \in E_m$ and $t \in [0, t^*(m, \zeta) \wedge t^*(m, \zeta')]$, one has

$$|w_m(\Phi_m(\zeta, t)) - w_m(\Phi_m(\zeta', t))| \leq [w_m]_1 |\zeta - \zeta'|.$$

Let $[w]_1 = \max_{m \in M} [w_m]_1$.

Definition 2.6. For all $u \geq 0$, denote by $L_c^u(E)$ the set of functions $w \in B(E)$ Lipschitz continuous along the flow until time u ; i.e., the real-valued, bounded, measurable functions defined on E and satisfying the following conditions:

- For all $x \in E$, the map $w(\Phi(x, \cdot)) : [0, t^*(x) \wedge u] \rightarrow \mathbb{R}$ is continuous. If $t^*(x) \leq u$, then $\lim_{t \rightarrow t^*(x)} w(\Phi(x, t))$ exists and is denoted by $w(\Phi(x, t^*(x)))$.
- There exists $[w]_1^{E,u} \in \mathbb{R}^+$ such that for all $x, y \in E$ and $t \in [0, t^*(x) \wedge t^*(y) \wedge u]$, one has

$$|w(\Phi(x, t)) - w(\Phi(y, t))| \leq [w]_1^{E,u} |x - y|.$$

- There exists $[w]_2^{E,u} \in \mathbb{R}^+$ such that for all $x \in E$ and $t, t' \in [0, t^*(x) \wedge u]$, one has

$$|w(\Phi(x, t)) - w(\Phi(x, t'))| \leq [w]_2^{E,u} |t - t'|.$$

- There exists $[w]_*^{E,u} \in \mathbb{R}^+$ such that for all $x, y \in E$, if $t^*(x) \leq u$ and $t^*(y) \leq u$, one has

$$|w(\Phi(x, t^*(x))) - w(\Phi(y, t^*(y)))| \leq [w]_*^{E,u} |x - y|.$$

Remark 2.7. For all $u \leq u'$, one has $L_c^{u'}(E) \subset L_c^u(E)$ with $[w]_i^{E,u} \leq [w]_i^{E,u'}$ where $i \in \{1, 2, *\}$.

Remark 2.8. Definitions 2.3 and 2.6 correspond respectively to the Lipschitz and local Lipschitz continuity along the flow that is, along the trajectories of the process. They can be replaced by (local) Lipschitz assumptions on the flow Φ , t^* and w in the classical sense.

We will require the following assumptions.

Assumption 2.9. The jump rate λ is bounded and there exists $[\lambda]_1 \in \mathbb{R}^+$ such that for all $x, y \in E$ and $t \in [0, t^*(x) \wedge t^*(y)]$, one has

$$|\lambda(\Phi(x, t)) - \lambda(\Phi(y, t))| \leq [\lambda]_1 |x - y|.$$

Assumption 2.10. The deterministic exit time from E , denoted by t^* , is assumed to be bounded and Lipschitz continuous on E .

Remark 2.11. Since the deterministic exit time t^* is bounded by C_{t^*} , one may notice that $L_c^u(E)$ for $u \geq C_{t^*}$ is no other than $L_c(E)$.

Remark 2.12. In most practical applications, the physical properties of the system ensure that either t^* is bounded, or the problem has a natural finite deterministic time horizon t_f . In the latter case, there is no loss of generality in considering that t^* is bounded by this deterministic time horizon. This leads to replacing C_{t^*} by t_f . An example of such a situation is presented in an industrial example in Section 6.2.

Assumption 2.13. The Markov kernel Q is Lipschitz in the following sense: there exists $[Q] \in \mathbb{R}^+$ such that for all $u \geq 0$ and for all function $w \in L_c^u(E)$, one has

(1) for all $x, y \in E$ and $t \in [0, t^*(x) \wedge t^*(y) \wedge u]$,

$$|Qw(\Phi(x, t)) - Qw(\Phi(y, t))| \leq [Q][w]_1^{E, u} |x - y|.$$

(2) for all $x, y \in E$ such that $t^*(x) \vee t^*(y) \leq u$,

$$|Qw(\Phi(x, t^*(x))) - Qw(\Phi(y, t^*(y)))| \leq [Q]([w]_*^{E, u} + [w]_1^{E, u}) |x - y|.$$

Remark 2.14. Assumption 2.13 is slightly more restrictive than its counterpart in [6] (Assumption 2.5), because of the introduction of the state space $L_c^u(E)$. This is to ensure that the time-augmented process still satisfies a similar assumption; see Section 5.1.

3. Expectation

From now on, we will assume that $Z_0 = x$ a.s. for some $x \in E$. For all fixed $N \in \mathbb{N}^*$, we intend to numerically approximate the quantity

$$J_N(l, c)(x) = \mathbf{E}_x \left[\int_0^{T_N} l(X_t) dt + \sum_{j=1}^N c(X_{T_j^-}) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right], \quad (2)$$

where $l \in B(E)$, $c \in B(\partial E)$ and X_{t-} is the left limit of X_t . Thus, $X_{T_j^-}$ is the j -th prejump location. Since the boundary jumps occur exactly at the deterministic exit times from E , one has,

$$J_N(l, c)(x) = \mathbf{E}_x \left[\int_0^{T_N} l(X_t) dt + \sum_{j=1}^N c(\Phi(Z_{j-1}, t^*(Z_{j-1}))) \mathbb{1}_{\{S_j = t^*(Z_{j-1})\}} \right].$$

In many applications, $J_N(l, c)(x)$ appears as a cost or a reward function. The first term, that depends on l , is called the running cost and the second one, that depends on c , is the boundary jump cost.

The rest of this section is devoted to formulating the expectation above in a way that will allow us to derive a numerical computation method. The Lipschitz continuity property will be a crucial point when it comes to proving the convergence of our approximation scheme. For this reason, the first step of our approximation is to replace the indicator function in $J_N(l, c)(x)$ by a Lipschitz continuous

function. Then we present a recursive method yielding the required expectation. This recursive formulation will be the basis of our numerical method.

3.1. Lipschitz continuity. We introduce a regularity assumption on l and c .

Assumption 3.1. We assume that $l \in \mathbf{L}_c(E)$ and $c \in \mathbf{L}_c(\partial E)$.

Moreover, we replace the indicator function in $J_N(l, c)(x)$ by a Lipschitz continuous function δ^A , with $A > 0$. Let then

$$J_N^A(l, c)(x) = \mathbf{E}_x \left[\int_0^{T_N} l(X_t) dt + \sum_{j=1}^N c(\Phi(Z_{j-1}, t^*(Z_{j-1}))) \delta^A(Z_{j-1}, S_j) \right],$$

where δ^A is a triangular approximation of the indicator function. It is defined on $E \times \mathbb{R}$ by

$$\delta^A(x, t) = \begin{cases} A \left(t - \left(t^*(x) - \frac{1}{A} \right) \right) & \text{for } t \in \left[t^*(x) - \frac{1}{A}; t^*(x) \right], \\ -A \left(t - \left(t^*(x) + \frac{1}{A} \right) \right) & \text{for } t \in \left[t^*(x); t^*(x) + \frac{1}{A} \right], \\ 0 & \text{otherwise.} \end{cases}$$

For all $x \in E$, the function $\delta^A(x, t)$ goes to $\mathbb{1}_{\{t=t^*(x)\}}$ when A goes to infinity. The following proposition proves the convergence of $J_N^A(l, c)(x)$ towards $J_N(l, c)(x)$ with an error bound.

Proposition 3.2. For all $x \in E$, $A > 0$, $N \in \mathbb{N}^*$, $l \in \mathbf{L}_c(E)$ and $c \in \mathbf{L}_c(\partial E)$, one has

$$|J_N^A(l, c)(x) - J_N(l, c)(x)| \leq \frac{NC_c C_\lambda}{A}.$$

Proof. For all $x \in E$, one has

$$\begin{aligned} & |J_N^A(l, c)(x) - J_N(l, c)(x)| \\ &= \left| \mathbf{E}_x \left[\sum_{j=1}^N c(\Phi(Z_{j-1}, t^*(Z_{j-1}))) (\delta^A(Z_{j-1}, S_j) - \mathbb{1}_{\{S_j=t^*(Z_{j-1})\}}) \right] \right| \\ &\leq C_c \sum_{j=1}^N \mathbf{E}_x [|\delta^A(Z_{j-1}, S_j) - \mathbb{1}_{\{S_j=t^*(Z_{j-1})\}}|] \\ &\leq C_c \sum_{j=1}^N \mathbf{E}_x [\mathbf{E} [|\delta^A(Z_{j-1}, S_j) - \mathbb{1}_{\{S_j=t^*(Z_{j-1})\}}| \mid Z_{j-1}]]. \end{aligned}$$

We recall that the conditional law of S_j with respect to Z_{j-1} has density

$$s \rightarrow \lambda(\Phi(Z_{j-1}, s)) e^{-\Lambda(Z_{j-1}, s)}$$

on $[0; t^*(Z_{j-1}))$ and puts the weight $e^{-\Lambda(Z_{j-1}, t^*(Z_{j-1}))}$ on the point $t^*(Z_{j-1})$. We also recall that λ is bounded thanks to Assumption 2.9. Finally, one has

$$\begin{aligned} & |J_N^A(l, c)(x) - J_N(l, c)(x)| \\ & \leq C_c \sum_{j=1}^N \mathbf{E}_x \left[\int_{t^*(Z_{j-1}) - \frac{1}{A}}^{t^*(Z_{j-1})} \delta^A(Z_{j-1}, s) \lambda(\Phi(Z_{j-1}, s)) e^{-\Lambda(Z_{j-1}, s)} ds \right] \\ & \leq \frac{NC_c C_\lambda}{A}. \end{aligned}$$

Hence the result. \square

Consequently to this proposition, we consider, from now on, the approximation of $J_N^A(l, c)(x)$ for some fixed A , large enough to ensure that the previous error is as small as required. The suitable choice of A will be discussed in Section 4.2.

3.2. Recursive formulation. Davis shows in [5, Section 32] that the expectation $J_N^A(l, c)(x)$ we are interested in is obtained by merely iterating an operator that we will denote by G . The rest of this section is dedicated to presenting this method from which we will derive our approximation scheme in Section 4.

Definition 3.3. Introduce the functions L , C and F defined for all $x \in E$ and $t \in [0; t^*(x)]$ by

$$\begin{aligned} L(x, t) &= \int_0^t l(\Phi(x, s)) ds, \\ C(x, t) &= c(\Phi(x, t^*(x))) \delta^A(x, t), \\ F(x, t) &= L(x, t) + C(x, t), \end{aligned}$$

along with the operator $G: B(E) \rightarrow B(E)$ given by

$$Gw(x) = \mathbf{E}_x[F(x, S_1) + w(Z_1)].$$

Definition 3.4. Define the sequence of functions $(v_k)_{0 \leq k \leq N}$ in $B(E)$ by

$$v_N(x) = 0, \quad v_k(x) = Gv_{k+1}(x).$$

Davis then shows in [5, Equation 32.33] that, for all $k \in \{0, \dots, N\}$,

$$v_{N-k}(x) = \mathbf{E}_x \left[\int_0^{T_k} l(X_t) dt + \sum_{j=1}^k c(\Phi(Z_{j-1}, t^*(Z_{j-1}))) \delta^A(Z_{j-1}, S_j) \right].$$

Thus, the quantity $J_N^A(l, c)(x)$ we intend to approximate is none other than $v_0(x)$.

Notice that, thanks to the Markov property of the chain $(Z_n, S_n)_{n \in \mathbb{N}}$, one has for all $k \in \{0, \dots, N-1\}$,

$$Gw(x) = E[F(Z_k, S_{k+1}) + w(Z_{k+1}) | Z_k = x]. \quad (3)$$

Hence, for all $k \in \{0, \dots, N\}$, let $V_k = v_k(Z_k)$ so that one has

$$V_N = 0, \quad V_k = E[F(Z_k, S_{k+1}) + V_{k+1} | Z_k].$$

This backward recursion provides the required quantity

$$V_0 = J_N^A(l, c)(x).$$

Hence, we need to approximate the sequence of random variables $(V_k)_{0 \leq k \leq N}$. This sequence satisfies a recursion that only depends on the chain $(Z_k, S_k)_{0 \leq k \leq N}$. Therefore, it appears natural to propose an approximation scheme based on a discretization of this chain $(Z_k, S_k)_{0 \leq k \leq N}$, called quantization, similarly to the ideas developed in [6] and [3].

4. Approximation scheme

Let us now turn to the approximation scheme itself. We explained in the previous section how the expectation we are interested in stems from the iteration of the operator G that only depends on the discrete-time Markov chain $(Z_k, S_k)_{0 \leq k \leq N}$. The first step of our numerical method is therefore to discretize this chain in order to approximate the operator G .

4.1. Quantization of the chain $(Z_k, S_k)_{k \leq N}$. Our approximation method is based on the quantization of the underlying discrete time Markov chain $(\Theta_k)_{k \leq N} = (Z_k, S_k)_{k \leq N}$. This quantization consists in finding an optimally designed discretization of the process to provide for each step k the best possible approximation of Θ_k by a random variable $\hat{\Theta}_k$ which state space has a finite and fixed number of points. Here, *optimal* means that the distance between Θ_k and $\hat{\Theta}_k$ in a suitably chosen L^p norm is minimal. For details on the quantization methods, we mainly refer to [9] but the interested reader can also consult [1], [2] and the references therein.

More precisely, consider X an \mathbb{R}^q -valued random variable such that $\|X\|_p < \infty$ and let K be a fixed integer. The optimal L_p -quantization of the random variable X consists in finding the best possible L_p -approximation of X by a random vector $\hat{X} \in \{x^1, \dots, x^K\}$ taking at most K values: This procedure consists of two steps:

- (1) Find a finite weighted grid $\Gamma \subset \mathbb{R}^q$ with $\Gamma = \{x^1, \dots, x^K\}$.
- (2) Set $\hat{X} = \hat{X}^\Gamma$ where $\hat{X}^\Gamma = \text{proj}_\Gamma(X)$ with proj_Γ denotes the closest neighbor projection on Γ .

The asymptotic properties of the L_p -quantization are given by the following result; see [9], for example.

Theorem 4.1. *If $\mathbb{E}[|X|^{p+\eta}] < +\infty$ for some $\eta > 0$ then one has*

$$\lim_{K \rightarrow \infty} K^{p/q} \min_{|\Gamma| \leq K} \|X - \hat{X}^\Gamma\|_p^p = J_{p,q} \int |h|^{q/(q+p)}(u) du,$$

where the law of X is $P_X(du) = h(u)\lambda_q(du) + \nu$ with $\nu \perp \lambda_q$, $J_{p,q}$ a constant and λ_q the Lebesgue measure in \mathbb{R}^q .

Remark that X needs to have finite moments up to the order $p + \eta$ to ensure the above convergence. In this work, we used the CLVQ quantization algorithm described in [1], Section 3.

There exists a similar procedure for the optimal quantization of a Markov chain $\{X_k\}_{k \in \mathbb{N}}$. There are two approaches to provide the quantized approximation of a Markov chain. The first one, based on the quantization at each time k of the random variable X_k is called the *marginal quantization*. The second one that enhances the preservation of the Markov property is called *Markovian quantization*. Remark that for the latter, the quantized Markov process is not homogeneous. These two methods are described in details in [9, Section 3]. In this work, we used the marginal quantization approach for simplicity reasons.

The quantization algorithm provides for each time step $0 \leq k \leq N$ a finite grid Γ_k of $E \times \mathbb{R}^+$ as well as the transition matrices $(\hat{Q}_k)_{0 \leq k \leq N-1}$ from Γ_k to Γ_{k+1} . Let $p \geq 1$ such that for all $k \leq N$, Z_k and S_k have finite moments at least up to order p and let proj_{Γ_k} be the closest-neighbor projection from $E \times \mathbb{R}^+$ onto Γ_k (for the distance associated to norm p). The quantized process $(\hat{\Theta}_k)_{k \leq N} = (\hat{Z}_k, \hat{S}_k)_{k \leq N}$ takes values for each k in the finite grid Γ_k of $E \times \mathbb{R}^+$ and is defined by

$$(\hat{Z}_k, \hat{S}_k) = \text{proj}_{\Gamma_k}(Z_k, S_k). \quad (4)$$

Moreover, we also denote by Γ_k^Z and Γ_k^S , respectively, the projections of Γ_k on E and \mathbb{R}^+ .

Some important remarks must be made concerning the quantization. On the one hand, the optimal quantization has nice convergence properties stated by Theorem 4.1. Indeed, the L^p -quantization error $\|\Theta_k - \hat{\Theta}_k\|_p$ goes to zero when the number of points in the grids goes to infinity. However, on the other hand, the Markov property is not maintained by the algorithm and the quantized process is generally not Markovian. Although the quantized process can be easily transformed into a Markov chain (see [9]), this chain will not be homogeneous. It must be pointed out that the quantized process $(\hat{\Theta}_k)_{k \in \mathbb{N}}$ depends on the starting point Θ_0 of the process.

In practice, we begin with the computation of the quantization grids which merely requires to be able to simulate the process. This step is quite time-consuming, especially when the number of points in the quantization grids is large. However, the grids are only computed once and for all and may be stored off-line. What is more, they only depend on the dynamics of the process, not on the cost functions l and c . Hence, the same grids may be used to compute different expectations of functionals as long as they are related to the same process. Our schemes are then based on the following simple idea: we replace the process by its quantized approximation within the operator G . The approximation is thus obtained in a very simple way since the quantized process has finite state space.

4.2. Approximation of the expectation and rate of convergence. We now use the quantization of the process $(\Theta_k)_{k \leq N} = (Z_k, S_k)_{k \leq N}$. In order to approximate the random variables $(V_k)_{k \leq N}$, we introduce a quantized version of the operator G . Notice that the quantized process is no longer an homogeneous Markov chain so that we have different operators for each time step k . Their definitions naturally stem from a remark made in the previous section: recall that for all $k \in \{1, \dots, N\}$ and $x \in E$,

$$Gw(x) = E[F(Z_{k-1}, S_k) + w(Z_k) \mid Z_{k-1} = x].$$

Definition 4.2. For all $k \in \{1, \dots, N\}$, $w \in B(\Gamma_k^Z)$ and $z \in \Gamma_{k-1}^Z$, let

$$\hat{G}_k w(z) = E[F(z, \hat{S}_k) + w(\hat{Z}_k) \mid \hat{Z}_{k-1} = z].$$

Introduce also the functions $(\hat{v}_k)_{0 \leq k \leq N}$ by

$$\begin{aligned} \hat{v}_N(z) &= 0 && \text{for all } z \in \Gamma_N^Z, \\ \hat{v}_k(z) &= \hat{G}_{k+1} \hat{v}_{k+1}(z) && \text{for all } k \in \{0, \dots, N-1\} \text{ and } z \in \Gamma_k^Z. \end{aligned}$$

Finally, for all $k \in \{0, \dots, N\}$, let

$$\hat{V}_k = \hat{v}_k(\hat{Z}_k).$$

Remark 4.3. The conditional expectation in $\hat{G}_k w(z)$ is a finite sum. Thus, the numerical computation of the sequence $(\hat{V}_k)_k$ will be easily performed as soon as the quantized process $(\hat{\Theta}_k)_{k \leq N}$ has been obtained.

Remark 4.4. We have assumed that $Z_0 = x$ a.s. Thus, the quantization algorithm provides that $\hat{Z}_0 = x$ a.s. too. Consequently, the random variable $\hat{V}_0 = \hat{v}_0(\hat{Z}_0)$ is, in fact, deterministic.

The following theorem states the convergence of \hat{V}_0 towards $V_0 = J_N^A(l, c)(x)$ and provides a bound for the rate of convergence.

Theorem 4.5. *For all $k \in \{0, \dots, N\}$, one has $v_k \in \mathbf{L}_c(E)$. Moreover, the approximation error satisfies*

$$|J_N(l, c)(x) - \widehat{V}_0| \leq \varepsilon_N(l, c, X, A),$$

where

$$\begin{aligned} \varepsilon_N(l, c, X, A) = \sum_{k=0}^{N-1} & \left(2[v_{k+1}] \|Z_{k+1} - \widehat{Z}_{k+1}\|_p + (2[v_k] + [F]_1) \|Z_k - \widehat{Z}_k\|_p \right. \\ & \left. + [F]_2 \|S_{k+1} - \widehat{S}_{k+1}\|_p \right) + \frac{NC_c C_\lambda}{A} \end{aligned}$$

with

$$[F]_1 = C_{t^*}[l]_1 + [c]_* + A[t]_* C_c,$$

$$[F]_2 = C_l + AC_c.$$

$$C_{v_n} \leq n(C_{t^*}C_l + C_c),$$

$$[v_n]_1 \leq e^{C_{t^*}C_\lambda} (K(A, v_{n-1}) + nC_{t^*}[\lambda]_1(C_{t^*}C_l + C_c)) + C_{t^*}[l]_1,$$

$$[v_n]_2 \leq e^{C_{t^*}C_\lambda} (C_{t^*}C_l C_\lambda + 2C_l + C_\lambda C_c + (2n-1)C_\lambda(C_{t^*}C_l + C_c)) + C_l,$$

$$[v_n]_* \leq [v_n]_1 + [t^*][v_n]_2,$$

$$[v_n] \leq K(A, v_{n-1}),$$

and for all $w \in \mathbf{L}_c(E)$, $K(A, w) = E_1 + E_2 A + E_3[w]_1 + E_4 C_w + [Q][w]_*$, where

$$\begin{aligned} E_1 = 2[l]_1 C_{t^*} + C_l([t^*] + 2C_{t^*}^2[\lambda]_1) + [c]_*(1 + C_{t^*}C_\lambda) \\ + C_c(2[\lambda]_1 C_{t^*} + C_\lambda C_{t^*}^2[\lambda]_1 + 2[t^*]C_\lambda), \end{aligned}$$

$$E_2 = C_c C_{t^*} C_\lambda [t^*],$$

$$E_3 = (1 + C_{t^*}C_\lambda)[Q],$$

$$E_4 = 2C_\lambda [t^*] + C_{t^*}[\lambda]_1(2 + C_{t^*}C_\lambda).$$

The choice of A . Proposition 3.2 suggests that A should be as large as possible. However, the constants $[F]_1$, $[F]_2$ and $[v_n]$ that appear in the bound of the approximation error proposed by the above Theorem 4.5 grow linearly with A . Thus, in order to control this error, it is necessary that the order of magnitude of the quantization error $\|\Theta_k - \widehat{\Theta}_k\|_p$ be at most $1/A$.

The convergence of the approximation scheme can be derived from Theorem 4.5. Indeed, on the one hand, one must remind that $V_0 = J_N^A(l, c)(x)$ is the expectation we intended to approximate and on the other hand, $\|\Theta_k - \widehat{\Theta}_k\|_p$ may become arbitrarily small when the number of points in the quantization grids goes to infinity (see [9], for example). An outline of the proof is presented in Appendix C.

5. Time-dependent functionals

We now turn to the main contribution of this paper and present two generalizations of the previous problem. On the one hand, we will consider time-dependent functionals of the form

$$E_x \left[\int_0^{T_N} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right]$$

where l and c are Lipschitz continuous functions. On the other hand, we wish to replace the random time horizon T_N by a deterministic one, denoted by t_f :

$$E_x \left[\int_0^{t_f} l(X_t, t) dt + \sum_{T_j \leq t_f} c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right].$$

We will reason as follows. As suggested by Davis in [5], we will introduce a transformation $(\tilde{X}_t)_{t \geq 0}$ of the initial process $(X_t)_{t \geq 0}$ by including the time variable into the state space: $(\tilde{X}_t) = (X_t, t)$. Indeed, we will see that both the expectation of the time-dependent functional and the one with deterministic time horizon are no other than expectations of time invariant functionals for the time-augmented process $(\tilde{X}_t)_{t \geq 0}$. We therefore intend to apply the previously exposed approximation scheme to this new PDMP. However, it is far from obvious that the Lipschitz continuity assumptions 2.9, 2.13 and 2.10 still hold for this new process.

Thus, the rest of this section is organized as follows. First, we recall the precise definition of the time-augmented process and prove that it satisfies the Lipschitz continuity assumptions required by our approximation scheme. Then, we will see that the time-dependent functional case corresponds to a time invariant functional for the new transformed process and may therefore be obtained thanks to the earlier method. Finally, we consider the deterministic time horizon problem that features an additional hurdle namely the presence of non-Lipschitz continuous indicator functions.

5.1. The time-augmented process. Davis suggests, in [5, Section 31], that the case of the time-dependent functionals may be treated by introducing the time variable within the state space. Thus, it will be possible to apply our previous numerical method to the time-augmented process. However, and this is what we discuss in this section, it is necessary to check whether the Lipschitz continuity assumptions still hold. We first recall the definition of the time-augmented process given by Davis.

Definition 5.1. Introduce the new state space

$$\tilde{E} = E \times \mathbb{R}^+$$

equipped with the norm defined by: for all $\xi = (x, t)$, $\xi' = (x', t') \in \tilde{E}$, let

$$|\xi - \xi'| = |x - x'| + |t - t'| \quad (5)$$

where the norm on E is given by (1). On this state space, we define the process

$$\tilde{X}_t = (X_t, t).$$

The local characteristics of the PDMP $(\tilde{X}_t)_{t \geq 0}$, denoted by $(\tilde{\lambda}, \tilde{Q}, \tilde{\Phi})$, are given for all $\xi = (x, t) \in \tilde{E}$ by

$$\begin{cases} \tilde{\lambda}(\xi) = \lambda(x), \\ \tilde{\Phi}(\xi, s) = (\Phi(x, s), t + s) & \text{for } s \leq t^*(x), \\ \tilde{Q}(\xi, A \times \{t\}) = Q(x, A) & \text{for all } A \in \mathcal{B}(E). \end{cases}$$

Moreover, we naturally define for all $\xi = (x, t) \in \tilde{E}$

$$\tilde{t}^*(\xi) = \inf\{s > 0 \text{ such that } \tilde{\Phi}(\xi, s) \in \partial\tilde{E}\} = t^*(x)$$

Clearly, Assumptions 2.9 and 2.10 still hold with $[\tilde{\lambda}]_1 = [\lambda]_1$ and $[\tilde{t}^*] = [t^*]$. However, proving Assumption 2.13 is more intricate. We start with the following lemma.

Lemma 5.2. *Let $u, t \geq 0$ and $w \in L_c^u(\tilde{E})$. Denote by w_t the function of $B(E)$ defined by $w_t = w(\cdot, t)$. One has then $w_t \in L_c^{t \wedge u}(E)$ with*

$$\begin{aligned} [w_t]_1^{E, t \wedge u} &\leq [w]_1^{\tilde{E}, u}, \\ [w_t]_2^{E, t \wedge u} &\leq [w]_1^{\tilde{E}, u} + [w]_2^{\tilde{E}, u}, \\ [w_t]_*^{E, t \wedge u} &\leq (1 + [t^*])[w]_*^{\tilde{E}, u}. \end{aligned}$$

Proof. Let $u, t \geq 0$ and $w \in L_c^u(\tilde{E})$. For $x, x' \in E$ and $s \leq t^*(x) \wedge t^*(x') \wedge t \wedge u$, one has

$$|w_t(\Phi(x, s)) - w_t(\Phi(x', s))| = |w(\tilde{\Phi}((x, t - s), s)) - w(\tilde{\Phi}((x', t - s), s))|.$$

We now use the fact that $w \in L_c^u(\tilde{E})$ which yields since $s \leq u$

$$|w_t(\Phi(x, s)) - w_t(\Phi(x', s))| \leq [w]_1^{\tilde{E}, u} |(x, t - s) - (x', t - s)| = [w]_1^{\tilde{E}, u} |x - x'|.$$

Hence, $[w_t]_1^{E, t \wedge u} \leq [w]_1^{\tilde{E}, u}$, and similarly one obtains $[w_t]_2^{E, t \wedge u} \leq [w]_1^{\tilde{E}, u} + [w]_2^{\tilde{E}, u}$.

On the other hand, for $x, x' \in E$ such that $t^*(x) \vee t^*(x') \leq t \wedge u$, one has

$$\begin{aligned} & |w_t(\Phi(x, t^*(x))) - w_t(\Phi(x', t^*(x')))| \\ &= |w(\tilde{\Phi}((x, t - t^*(x)), t^*(x))) - w(\tilde{\Phi}((x', t - t^*(x')), t^*(x')))| \\ &= \left| w(\tilde{\Phi}((x, t - t^*(x)), \tilde{t}^*(x, t - t^*(x)))) \right. \\ & \quad \left. - w(\tilde{\Phi}((x', t - t^*(x')), \tilde{t}^*(x', t - t^*(x')))) \right|; \end{aligned}$$

moreover since $w \in L_c^u(\tilde{E})$ and $\tilde{t}^*(x, t - t^*(x)) \vee \tilde{t}^*(x', t - t^*(x')) \leq u$ one has

$$|w_t(\Phi(x, t^*(x))) - w_t(\Phi(x', t^*(x')))| \leq [w]_*^{\tilde{E}, u} |(x, t - t^*(x)) - (x', t - t^*(x'))|.$$

We conclude thanks to the Lipschitz continuity assumption 2.10 on t^* , which yields $|(x, t - t^*(x)) - (x', t - t^*(x'))| \leq (1 + [t^*])|x - x'|$. One obtains $[w_t]_*^{E, t \wedge u} \leq [w]_*^{\tilde{E}, u} (1 + [t^*])$ and $w_t \in L_c^{t \wedge u}(E)$. \square

The next proposition proves that Assumption 2.13 holds for the time-augmented process $(\tilde{X})_{t \geq 0}$.

Proposition 5.3. *Let $w \in L_c^u(\tilde{E})$.*

(1) *For all $\xi, \xi' \in \tilde{E}$ and $s \in [0, \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi') \wedge u]$,*

$$|\tilde{Q}w(\tilde{\Phi}(\xi, s)) - \tilde{Q}w(\tilde{\Phi}(\xi', s))| \leq ([Q] \vee 1)[w]_1^{\tilde{E}, u} |\xi - \xi'|.$$

(2) *For all $\xi, \xi' \in \tilde{E}$ such that $\tilde{t}^*(\xi) \vee \tilde{t}^*(\xi') \leq u$,*

$$\begin{aligned} & |\tilde{Q}w(\tilde{\Phi}(\xi, \tilde{t}^*(\xi))) - \tilde{Q}w(\tilde{\Phi}(\xi', \tilde{t}^*(\xi')))| \\ & \leq ([Q] \vee 1)(1 + [t^*])([w]_*^{\tilde{E}, u} + [w]_1^{\tilde{E}, u}) |\xi - \xi'|. \end{aligned}$$

In other words, Assumption 2.13 is satisfied with $[\tilde{Q}] = ([Q] \vee 1)(1 + [t^])$.*

Proof. As in the previous lemma, for all $t \geq 0$, we will denote by w_t the function of $B(E)$ defined by $w_t = w(\cdot, t)$. For $\xi = (x, t) \in \tilde{E}$ and $w \in L_c^u(\tilde{E})$, one has, by the definition of \tilde{Q} ,

$$\tilde{Q}w(\xi) = \int_{\xi' \in \tilde{E}} w(\xi') \tilde{Q}((x, t), d\xi') = \int_{z \in E} w(z, t) Q(x, dz) = Qw_t(x). \quad (6)$$

We may now check the regularity assumption on \tilde{Q} . Let $\xi = (x, t)$ and $\xi' = (x', t') \in \tilde{E}$. Let $s \in [0, \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi') \wedge u]$. Thanks to the definition of $\tilde{\Phi}$ and (6) one has

$$\begin{aligned} & |\tilde{Q}w(\tilde{\Phi}(\xi, s)) - \tilde{Q}w(\tilde{\Phi}(\xi', s))| = |\tilde{Q}w(\Phi(x, s), t + s) - \tilde{Q}w(\Phi(x', s), t' + s)| \\ & = |Qw_{t+s}(\Phi(x, s)) - Qw_{t'+s}(\Phi(x', s))|. \end{aligned}$$

We split this into the sum of two differences:

$$\begin{aligned} & |Qw_{t+s}(\Phi(x, s)) - Qw_{t'+s}(\Phi(x', s))| \\ & \leq |Qw_{t+s}(\Phi(x, s)) - Qw_{t+s}(\Phi(x', s))| + |Q(w_{t+s} - w_{t'+s})(\Phi(x', s))|. \end{aligned}$$

On the one hand, we recall that thanks to Lemma 5.2, $w_{t+s} \in \mathbf{L}_c^{(t+s) \wedge u}(E)$, so that, since $s \leq (t+s) \wedge u$, we may use the Lipschitz continuity assumption 2.13 on Q and the first term is bounded as follows:

$$|Qw_{t+s}(\Phi(x, s)) - Qw_{t+s}(\Phi(x', s))| \leq [Q][w_{t+s}]_1^{E, (t+s) \wedge u} |x - x'|.$$

Lemma 5.2 also provides $[w_{t+s}]_1^{E, (t+s) \wedge u} \leq [w]_1^{\tilde{E}, u}$. On the other hand, and more basically, the second term in the equation above satisfies

$$|Q(w_{t+s} - w_{t'+s})(\Phi(x', s))| \leq [w]_1^{\tilde{E}, u} |t - t'|.$$

We obtain

$$|\tilde{Q}w(\tilde{\Phi}(\xi, s)) - \tilde{Q}w(\tilde{\Phi}(\xi', s))| \leq ([Q] \vee 1)[w]_1^{\tilde{E}, u} |\xi - \xi'|.$$

We reason similarly to bound $|\tilde{Q}w(\tilde{\Phi}(\xi, \tilde{t}^*(\xi))) - \tilde{Q}w(\tilde{\Phi}(\xi', \tilde{t}^*(\xi')))|$, where $\xi = (x, t)$ and $\xi' = (x', t') \in \tilde{E}$ are such that $\tilde{t}^*(\xi) \vee \tilde{t}^*(\xi') \leq u$. Equation (6) yields

$$\begin{aligned} & |\tilde{Q}w(\tilde{\Phi}(\xi, \tilde{t}^*(\xi))) - \tilde{Q}w(\tilde{\Phi}(\xi', \tilde{t}^*(\xi')))| \\ & = |Qw_{t+\tilde{t}^*(x)}(\Phi(x, \tilde{t}^*(x))) - Qw_{t'+\tilde{t}^*(x')}(\Phi(x', \tilde{t}^*(x')))|, \end{aligned}$$

which we now split as follows:

$$\begin{aligned} & |Qw_{t+\tilde{t}^*(x)}(\Phi(x, \tilde{t}^*(x))) - Qw_{t'+\tilde{t}^*(x')}(\Phi(x', \tilde{t}^*(x')))| \\ & \leq |Qw_{t+\tilde{t}^*(x)}(\Phi(x, \tilde{t}^*(x))) - Qw_{t+\tilde{t}^*(x)}(\Phi(x', \tilde{t}^*(x')))| \\ & \quad + |(Qw_{t+\tilde{t}^*(x)} - Qw_{t'+\tilde{t}^*(x')})(\Phi(x', \tilde{t}^*(x')))|. \end{aligned}$$

Thanks to Lemma 5.2, $w_{t+\tilde{t}^*(x)} \in \mathbf{L}_c^{(t+\tilde{t}^*(x)) \wedge u}(E)$. We assume, without loss of generality, that $\tilde{t}^*(x) \geq \tilde{t}^*(x')$, so $\tilde{t}^*(x) \vee \tilde{t}^*(x') \leq (t + \tilde{t}^*(x)) \wedge u$. Therefore, the first term in the above equation is bounded, thanks to the Lipschitz continuity assumption 2.13 on Q and Lemma 5.2, by

$$[Q]((1 + [\tilde{t}^*])[w]_*^{\tilde{E}, u} + [w]_1^{\tilde{E}, u})|x - x'|.$$

It is more straightforward to obtain a bound for the second term, of the form

$$[w]_1^{\tilde{E}, u} |t - t' + \tilde{t}^*(x) - \tilde{t}^*(x')| \leq [w]_1^{\tilde{E}, u} (|t - t'| + [\tilde{t}^*]|x - x'|).$$

We obtain

$$\begin{aligned} & |\tilde{Q}w(\tilde{\Phi}(\xi, \tilde{t}^*(\xi))) - \tilde{Q}w(\tilde{\Phi}(\xi', \tilde{t}^*(\xi')))| \\ & \leq [Q](1 + [t^*])[w]_*^{\tilde{E}, u} |x - x'| + [w]_1^{\tilde{E}, u} ([Q]|x - x'| + |t - t'| + [t^*]|x - x'|) \\ & \leq ([Q] \vee 1)(1 + [t^*])([w]_*^{\tilde{E}, u} + [w]_1^{\tilde{E}, u})|\xi - \xi'|. \end{aligned}$$

Hence the result. \square

Consequently, we may apply our numerical method to the time-augmented process $(\tilde{X}_t)_{t \geq 0}$. In other words, for $l \in \mathbf{L}_c(\tilde{E})$, $c \in \mathbf{L}_c(\partial \tilde{E})$ and $\xi \in \tilde{E}$, our approximation scheme may be used to compute

$$\tilde{J}_N(l, c)(\xi) = \mathbf{E}_\xi \left[\int_0^{T_N} l(\tilde{X}_t) dt + \sum_{j=1}^N c(\tilde{X}_{T_j^-}) \mathbb{1}_{\{\tilde{X}_{T_j^-} \in \partial \tilde{E}\}} \right]. \quad (7)$$

We will now see that the time-dependent functional and the deterministic time horizon problems boil down to computing such quantities $\tilde{J}_N(l, c)(\xi)$ for suitably chosen functions l and c .

5.2. Lipschitz continuous cost functions. We first consider the time-dependent functional problem with Lipschitz continuous cost functions. Thus, let then $l \in \mathbf{L}_c(\tilde{E})$, $c \in \mathbf{L}_c(\partial \tilde{E})$ and $x \in E$, we wish to compute

$$\mathbf{E}_x \left[\int_0^{T_N} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right].$$

It is straightforward to show that this quantity may be expressed using the time-augmented process starting from the point $\xi_0 = (x, 0)$. Indeed, one has

$$\tilde{J}_N(l, c)(\xi_0) = \mathbf{E}_x \left[\int_0^{T_N} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right],$$

where $\tilde{J}_N(l, c)(\xi_0)$ is given by (7). Although they are time-dependent, the cost functions l and c are seen, in the left-hand side term, as time invariant functions of the time-augmented process. The expectation of the time-dependent functional is therefore obtained by computing the expectation of a time invariant functional for the transformed PDMP thanks to the approximation scheme described in Section 4. This is what expresses the following theorem, which proof stems from the previous discussion.

Theorem 5.4. *Let $l \in \mathbf{L}_c(\tilde{E})$ and $c \in \mathbf{L}_c(\partial \tilde{E})$ and apply the approximation scheme described in Section 4 to the time-augmented process $(\tilde{X}_t)_{t \geq 0}$, one has then*

$$\left| \mathbf{E}_x \left[\int_0^{T_N} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right] - \widehat{V}_0 \right| \leq \varepsilon_N(l, c, \tilde{X}, A),$$

where we denote by $\varepsilon_N(l, c, \tilde{X}, A)$ the bound of the approximation error provided by Theorem 4.5 when our approximation scheme is applied with cost functions l and c to the time-augmented process $(\tilde{X}_t)_{t \geq 0}$.

Remark 5.5. The quantity $\varepsilon_N(l, c, \tilde{X}, A)$ is computed with respect to the process $(\tilde{X}_t)_{t \geq 0}$ instead of $(X_t)_{t \geq 0}$, as presented in Theorem 4.5, so that

$$\begin{aligned} \varepsilon_N(l, c, \tilde{X}, A) = & \sum_{k=0}^{N-1} \left(2[v_{k+1}]^{\tilde{E}} \|\tilde{Z}_{k+1} - \widehat{Z}_{k+1}\|_p \right. \\ & + (2[v_k]^{\tilde{E}} + [F]_1' + [F]_1'' A) \|\tilde{Z}_k - \widehat{Z}_k\|_p \\ & \left. + ([F]_2' + A[F]_2'') \|\tilde{S}_{k+1} - \widehat{S}_{k+1}\|_p \right) + \frac{NC_c C_\lambda}{A}. \end{aligned}$$

where $(\tilde{Z}_k, \tilde{S}_k)_{k \in \mathbb{N}}$ denotes the sequence of the postjump locations and the inter-jump times of the time-augmented process $(\tilde{X}_t)_{t \geq 0}$, and where

$$\begin{aligned} [F]_1' &= C_{t^*} [l]_1^{\tilde{E}} + [c]_*^{\tilde{E}}, \\ [F]_1'' &= [t^*] C_c, \\ [F]_2' &= C_l, \\ [F]_2'' &= C_c, \\ C_{v_n} &\leq n(C_{t^*} C_l + C_c), \\ [v_n]_1^{\tilde{E}} &\leq e^{C_{t^*} C_\lambda} (\tilde{K}(A, v_{n-1}) + n C_{t^*} [\lambda]_1 (C_{t^*} C_l + C_c)) + C_{t^*} [l]_1^{\tilde{E}}, \\ [v_n]_2^{\tilde{E}} &\leq e^{C_{t^*} C_\lambda} (C_{t^*} C_l C_\lambda + 2C_l + C_\lambda C_c + (2n-1) C_\lambda (C_{t^*} C_l + C_c)) + C_l, \\ [v_n]_*^{\tilde{E}} &\leq [v_n]_1^{\tilde{E}} + [t^*] [v_n]_2^{\tilde{E}}, \\ [v_n]^{\tilde{E}} &\leq \tilde{K}(A, v_{n-1}), \end{aligned}$$

and for all $w \in L_c(E)$ we have

$$\tilde{K}(A, w) = \tilde{E}_1 + E_2 A + \tilde{E}_3 [w]_1^{\tilde{E}} + E_4 C_w + [\tilde{Q}] [w]_*^{\tilde{E}},$$

where

$$\begin{aligned} [\tilde{Q}] &= ([Q] \vee 1)(1 + [t^*]), \\ \tilde{E}_1 &= 2[l]_1^{\tilde{E}} C_{t^*} + C_l ([t^*] + 2C_{t^*}^2 [\lambda]_1) + [c]_*^{\tilde{E}} (1 + C_{t^*} C_\lambda) \\ &\quad + C_c (2[\lambda]_1 C_{t^*} + C_\lambda C_{t^*}^2 [\lambda]_1 + 2[t^*] C_\lambda), \\ E_2 &= C_c C_{t^*} C_\lambda [t^*], \end{aligned}$$

$$\begin{aligned}\tilde{E}_3 &= (1 + C_{t^*} C_\lambda)[\tilde{Q}], \\ E_4 &= 2C_\lambda[t^*] + C_{t^*}[\lambda]_1(2 + C_{t^*} C_\lambda).\end{aligned}$$

5.3. Deterministic time horizon. In the context of applications, it seems relevant to consider a deterministic time horizon t_f . For instance, one may want to estimate a mean cost over a given period no matter how many jumps occur during this period. Actually, we will choose a time horizon of the form $t_f \wedge T_N$ with N large enough to ensure the N -th jump will occur after time t_f with high probability: in other words, that $\mathbf{P}_x(T_N < t_f)$ be close to zero. For a discussion concerning the choice of such N , and in particular a theoretical bound of the probability $\mathbf{P}_x(T_N < t_f)$, we refer to [3]. Simply notice that in practice, this probability may be estimated through Monte Carlo simulations. We thus intend to approximate the following quantity for $l \in \mathbf{L}_c(\tilde{E})$, $c \in \mathbf{L}_c(\partial\tilde{E})$ and $x \in E$:

$$\begin{aligned}\mathbf{E}_x \left[\int_0^{T_N \wedge t_f} l(X_t, t) dt + \sum_{T_j \leq t_f} c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right] \\ = \mathbf{E}_x \left[\int_0^{T_N} l(X_t, t) \mathbb{1}_{\{t \leq t_f\}} dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} \right].\end{aligned}$$

The natural approach would consist in killing the process at time t_f as Davis suggests in [5, Section 31], and applying our method to the new process. However, the killed process will not necessarily fulfill our Lipschitz continuity assumptions because of the discontinuity introduced at time t_f .

A second idea would then be to use the previous results, to consider the time-augmented process, and to define $\tilde{l}(x, t) = l(x, t) \mathbb{1}_{\{t \leq t_f\}}$ and $\tilde{c}(x, t) = c(x, t) \mathbb{1}_{\{t \leq t_f\}}$. However, a similar problem appears. Indeed, such functions \tilde{l} and \tilde{c} are not Lipschitz continuous and our numerical method requires this assumption. In the rest of this section, we will see how to overcome this drawback. On the one hand, we prove that the Lipschitz continuity condition on l may be relaxed so that our numerical method may be used directly to approximate $\tilde{J}_N(\tilde{l}, c)$ for any $c \in \mathbf{L}_c(\partial\tilde{E})$. On the other hand, in the general case, we will deal with the non-Lipschitz continuity of \tilde{c} by bounding it between two Lipschitz continuous functions.

5.3.1. Direct estimation of the running cost term. Let us explain how the Lipschitz continuity condition on the running cost function may be relaxed so that Theorem 4.5, stating the convergence of our approximation scheme, remains true when the running cost function is $\tilde{l}(x, t) = l(x, t) \mathbb{1}_{\{t \leq t_f\}}$ with $l \in \mathbf{L}_c(\tilde{E})$ and the boundary jump cost function is $c \in \mathbf{L}_c(\partial\tilde{E})$ (although with slightly different constants in the bound of the convergence rate). Indeed, the running cost function \tilde{l} appears inside an integral that will have a regularizing effect allowing us to derive

the required Lipschitz property of the functional in spite of the discontinuity of \tilde{l} . Details are provided in Appendix B.

Consequently, our approximation scheme may be used to compute $\tilde{J}_N(\tilde{l}, c)(\xi)$ for any $\xi \in \tilde{E}$. We recall that \tilde{J}_N is defined by (7) and that for all $x \in E$, one has

$$\tilde{J}_N(\tilde{l}, c)(x, 0) = E_x \left[\int_0^{T_N \wedge t_f} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right].$$

We now turn to the indicator function $\mathbb{1}_{\{T_j \leq t_f\}}$ required within the boundary jump cost term.

5.3.2. Bounds of the boundary jump cost term. We explained how the Lipschitz continuity condition on l may be relaxed. However, when it comes to c , this condition cannot be avoided and our numerical method cannot be used directly with $\tilde{c}(x, t) = c(x, t) \mathbb{1}_{\{t \leq t_f\}}$. We overcome this drawback by using Lipschitz continuous approximations of the indicator function. Indeed, for $B > 0$, we introduce the real-valued functions \underline{u}_B and \bar{u}_B defined on \mathbb{R} by

$$\underline{u}_B(t) = \begin{cases} 1 & \text{if } t < t_f - 1/B, \\ -B(t - t_f) & \text{if } t_f - 1/B \leq t < t_f, \\ 0 & \text{if } t_f \leq t, \end{cases}$$

$$\bar{u}_B(t) = \begin{cases} 1 & \text{if } t < t_f, \\ -B(t - t_f) + 1 & \text{if } t_f \leq t < t_f + 1/B, \\ 0 & \text{if } t_f + 1/B \leq t. \end{cases}$$

The following lemma is straightforward.

Lemma 5.6. *For all $t \geq 0$, $\lim_{B \rightarrow +\infty} \underline{u}_B(t) = \mathbb{1}_{[0; t_f)}(t)$ and $\lim_{B \rightarrow +\infty} \bar{u}_B(t) = \mathbb{1}_{[0; t_f]}(t)$. Furthermore, for all $B > 0$, \underline{u}_B and \bar{u}_B are Lipschitz continuous with Lipschitz constant B . Moreover, $|\underline{u}_B - \mathbb{1}_{[0; t_f)}| \leq 1$, $|\bar{u}_B - \mathbb{1}_{[0; t_f]}| \leq 1$ and*

$$\underline{u}_B \leq \mathbb{1}_{[0; t_f)} \leq \bar{u}_B.$$

Thus, define for $l \in \mathbf{L}_c(\tilde{E})$

$$\tilde{l}(x, t) = l(x, t) \mathbb{1}_{\{t \leq t_f\}} \quad (8)$$

and for $c \in \mathbf{L}_c(\partial \tilde{E})$ and for all $B > 0$,

$$\underline{c}_B(x, t) = c(x, t) \underline{u}_B(t) \quad \text{and} \quad \bar{c}_B(x, t) = c(x, t) \bar{u}_B(t). \quad (9)$$

We now check that these functions satisfy our Lipschitz continuity conditions.

Proposition 5.7. *The functions \underline{c}_B and \bar{c}_B belong to $\mathbf{L}_c(\partial \tilde{E})$ with $[\underline{c}_B]_*$, $[\bar{c}_B]_* \leq [c]_* + BC_c(1 \vee [t^*])$.*

Proof. We prove the result for \underline{c}_B , the other case being similar. For all $\xi = (x, t), \xi' = (x', t') \in \tilde{E}$, one has

$$\begin{aligned} & |\underline{c}_B(\tilde{\Phi}(\xi, t^*(\xi))) - \underline{c}_B(\tilde{\Phi}(\xi', t^*(\xi')))| \\ &= |c(\tilde{\Phi}(\xi, t^*(\xi)))\underline{u}_B(t + t^*(\xi)) - c(\tilde{\Phi}(\xi', t^*(\xi')))\underline{u}_B(t' + t^*(\xi'))| \\ &\leq [c]_* |\xi - \xi'| + C_c |\underline{u}_B(t + t^*(\xi)) - \underline{u}_B(t' + t^*(\xi'))| \\ &\leq [c]_* |\xi - \xi'| + C_c B(|t - t'| + [t^*]|x - x'|) \\ &\leq ([c]_* + C_c B(1 \vee [t^*])) |\xi - \xi'|. \end{aligned}$$

Hence the result. \square

Therefore, the functions \underline{c}_B and \bar{c}_B are acceptable boundary jump cost functions and we may bound the deterministic horizon expectation by

$$\begin{aligned} \tilde{J}_N(\tilde{l}, \underline{c}_B)(x, 0) &\leq \mathbf{E}_x \left[\int_0^{T_N} l(X_t) \mathbb{1}_{\{t \leq t_f\}} dt + \sum_{j=1}^N c(X_{T_j^-}) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} \right] \\ &\leq \tilde{J}_N(\tilde{l}, \bar{c}_B)(x, 0). \end{aligned}$$

The following proposition provides the convergence of the bounds.

Proposition 5.8. *For all $x \in E$, one has*

$$\begin{aligned} & \lim_{B \rightarrow +\infty} \tilde{J}_N(\tilde{l}, \underline{c}_B)(x, 0) \\ &= \lim_{B \rightarrow +\infty} \tilde{J}_N(\tilde{l}, \bar{c}_B)(x, 0) \\ &= \mathbf{E}_x \left[\int_0^{T_N \wedge t_f} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} \right]. \end{aligned}$$

Convergence holds for every $t_f > 0$ in the case of $\tilde{J}_N(\tilde{l}, \bar{c}_B)(x, 0)$ but only for almost every $t_f > 0$ with respect to the Lebesgue measure on \mathbb{R} in the case of $\tilde{J}_N(\tilde{l}, \underline{c}_B)(x, 0)$.

Proof. Let $x \in E$. We first consider $\tilde{J}_N(\tilde{l}, \bar{c}_B)(x, 0)$.

$$\begin{aligned} & \left| \mathbf{E}_x \left[\sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} - \sum_{j=1}^N \bar{c}_B(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right] \right| \\ &\leq \mathbf{E}_x \left[\sum_{j=1}^N |c(X_{T_j^-}, T_j)| |\mathbb{1}_{\{T_j \leq t_f\}} - \bar{u}_B(T_j)| \right] \\ &\leq C_c \mathbf{E}_x \left[\sum_{j=1}^N \mathbb{1}_{\{t_f < T_j \leq t_f + \frac{1}{B}\}} \right] \leq C_c \sum_{j=1}^N \left(\varphi_j \left(t_f + \frac{1}{B} \right) - \varphi_j(t_f) \right), \end{aligned}$$

where φ_j is the distribution function of T_j . For all $j \leq N$, the summand in this last expression goes to 0 as $B \rightarrow +\infty$, since φ_j is right-continuous; this shows the required convergence.

We now turn to the case of $\tilde{J}_N(\tilde{l}, \underline{c}_B)(x, 0)$. Similar computations yield

$$\begin{aligned} \left| \mathbf{E}_x \left[\sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} - \sum_{j=1}^N \underline{c}_B(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \right] \right| \\ \leq C_c \sum_{j=1}^N \left(\varphi_j(t_f) - \varphi_j\left(t_f - \frac{1}{B}\right) \right). \end{aligned}$$

One cannot conclude as in the previous case, since φ_j need not be left-continuous. We therefore assume that t_f is not an atom of any of the laws of the random variables T_j . Then, for all $j \leq N$, the summand on the right-hand side tends to 0 as $B \rightarrow +\infty$, and the result follows. Indeed, the set of the atoms of T_j is at most countable, so the convergence holds for almost every t_f with respect to the Lebesgue measure on \mathbb{R} . \square

5.3.3. Bounds in the general case. The previous results show that the deterministic horizon expectation may be bounded by applying our numerical method with \tilde{l} and successively \underline{c}_B and \bar{c}_B . In other words, we have shown:

Theorem 5.9. *Let $l \in \mathbf{L}_c(\tilde{E})$ and $c \in \mathbf{L}_c(\partial \tilde{E})$. Let $(\underline{V}_{k,B})_{0 \leq k \leq N}$ (respectively $(\bar{V}_{k,B})_{0 \leq k \leq N}$) be the sequence of random variables $(V_k)_{0 \leq k \leq N}$ described in Section 4 when applying our approximation scheme to the time-augmented process $(\tilde{X}_t)_{t \geq 0}$ with cost functions \tilde{l} and \underline{c}_B (respectively \bar{c}_B) defined by (8) and (9). The bounds of the approximation error provided by Theorem 4.5 are respectively denoted by*

$$\varepsilon_N(l, \underline{c}_B, \tilde{X}, A, B) \quad \text{and} \quad \varepsilon_N(l, \bar{c}_B, \tilde{X}, A, B).$$

One has then

$$\begin{aligned} \underline{V}_{0,B} - \varepsilon_N(l, \underline{c}_B, \tilde{X}, A, B) \\ \leq \mathbf{E}_x \left[\int_0^{T_N \wedge t_f} l(X_t, t) dt + \sum_{j=1}^N c(X_{T_j^-}, T_j) \mathbb{1}_{\{X_{T_j^-} \in \partial E\}} \mathbb{1}_{\{T_j \leq t_f\}} \right] \\ \leq \bar{V}_{0,B} + \varepsilon_N(l, \bar{c}_B, \tilde{X}, A, B). \end{aligned}$$

Remark 5.10. In the previous theorem, the quantity $\varepsilon_N(l, \underline{c}_B, \tilde{X}, A, B)$ (and similarly $\varepsilon_N(l, \bar{c}_B, \tilde{X}, A, B)$) is computed with respect to the process $(\tilde{X}_t)_{t \geq 0}$ instead

of $(X_t)_{t \geq 0}$ as presented in Theorem 4.5 so that one has

$$\begin{aligned} & \varepsilon_N(l, \underline{c}_B, \tilde{X}, A, B) \\ &= \sum_{k=0}^{N-1} \left(2[v_{k+1}]^{\tilde{E}} \|\tilde{Z}_{k+1} - \hat{\tilde{Z}}_{k+1}\|_p \right. \\ & \quad \left. + (2[v_k]^{\tilde{E}} + [F]_1' + [F]_1'' A + [F]_1''' B) \|\tilde{Z}_k - \hat{\tilde{Z}}_k\|_p \right. \\ & \quad \left. + ([F]_2' + [F]_2'' A) \|\tilde{S}_{k+1} - \hat{\tilde{S}}_{k+1}\|_p \right) + \frac{NC_c C_\lambda}{A}. \end{aligned}$$

where $(\tilde{Z}_k, \tilde{S}_k)_{k \in \mathbb{N}}$ denotes the sequence of the postjump locations and the inter-jump times of the time-augmented process $(\tilde{X}_t)_{t \geq 0}$ and with

$$\begin{aligned} [F]_1''' &= C_c(1 \vee [t^*]), \\ [v_n]_1^{\tilde{E}} &\leq e^{C_{t^*} C_\lambda} (\tilde{K}(A, B, v_{n-1}) + n C_{t^*} [\lambda]_1 (C_{t^*} C_l + C_c)) + C_{t^*} [l]_1^{\tilde{E}}, \\ [v_n]^{\tilde{E}} &\leq \tilde{K}(A, B, v_{n-1}), \end{aligned}$$

and for all $w \in L_c(E)$ we have

$$\tilde{K}(A, B, w) = E_1' + E_2'' B + E_2 A + \tilde{E}_3 [w]_1^{\tilde{E}} + E_4 C_w + [\tilde{Q}][w]_*^{\tilde{E}},$$

where

$$\begin{aligned} E_1' &= 2[l]_1^{\tilde{E}} C_{t^*} + C_l ([t^*] + 2C_{t^*}^2 [\lambda]_1) + [c]_*^{\tilde{E}} (1 + C_{t^*} C_\lambda) \\ & \quad + C_c (2[\lambda]_1 C_{t^*} + C_\lambda C_{t^*}^2 [\lambda]_1 + 2[t^*] C_\lambda), \\ E_1'' &= C_c (1 \vee [t^*]) (1 + C_{t^*} C_\lambda) \end{aligned}$$

The other constants remain unchanged; see Remark 5.5 for their expressions.

Furthermore, it is important to stress the fact that applying twice our numerical method does not increase significantly the computing time. Indeed, the computation of the quantization grids is, by far, the most costly step. These grids, that only depend on the dynamics of the process, may be stored off-line and used for the approximation of both bounds.

The choice of B . We now discuss the choice of the parameter B , the discussion is quite similar to the one concerning the choice of A in Section 4.2. Proposition 5.8 suggests that B should be chosen as large as possible. However, choosing a large value for B will lead to large Lipschitz constants that will decrease the sharpness of the bounds $\varepsilon_N(l, \underline{c}_B, \tilde{X})$ and $\varepsilon_N(l, \bar{c}_B, \tilde{X})$ for the approximation error provided by Theorem 4.5. Indeed, it is easy to check that $[v_n]$ grows linearly with B (see the precise expressions of the Lipschitz constants above). Thus, in order to control the error proposed by Theorem 4.5, it is necessary that the order of magnitude of the quantization error $\|\Theta_n - \hat{\Theta}_n\|_p$ be at most $1/B$.

6. Numerical results

6.1. A repair workshop model. We now present a repair workshop model adapted from [5, Section 21].

In a factory, a machine produces goods which daily value is $r(x)$, where $x \in [0; 1]$ represents a parameter of evolution of the machine, a setting chosen by the operator. For instance, x may be some load or some pace imposed on the machine. This machine, initially working, may break down with an age-dependent hazard rate $\lambda(t)$ and is then sent to the workshop for repair. Besides, the factory’s management has decided that, whenever the machine has worked for a whole year without requiring repair, it is sent to the workshop for maintenance. The daily cost of maintenance is $q(x)$, while the daily cost of a repairs is $p(x)$, with reasonably $p(x) > q(x)$. We assume that after a repair or maintenance, both lasting a fixed time s , the machine is totally repaired and is not worn down.

We therefore consider three modes: the machine is working ($m = 1$), being repaired ($m = 2$), or undergoing maintenance ($m = 3$). The state of the process at time t will be denoted by $X_t = (m_t, \zeta_t, t)$, where ζ_t is the time since the last change of mode. (This component is required since the hazard rate λ is age-dependent.) The state space is $E = (\{1\} \times [0; 365] \times \mathbb{R}^+) \cup (\{2\} \times [0; s] \times \mathbb{R}^+) \cup (\{3\} \times [0; s] \times \mathbb{R}^+)$. In each mode, the flow is $\Phi_m((\zeta, t), u) = (\zeta + u, t + u)$. Concerning the transition kernel, one sees from the previous discussion that, for instance, from the point $(1, \zeta, t)$, the process can jump to the point $(2, 0, t)$ if $\zeta < 365$ and the jump is forced to $(3, 0, t)$ if $\zeta = 365$. Figure 1 presents the state space and an example of trajectory of the process.

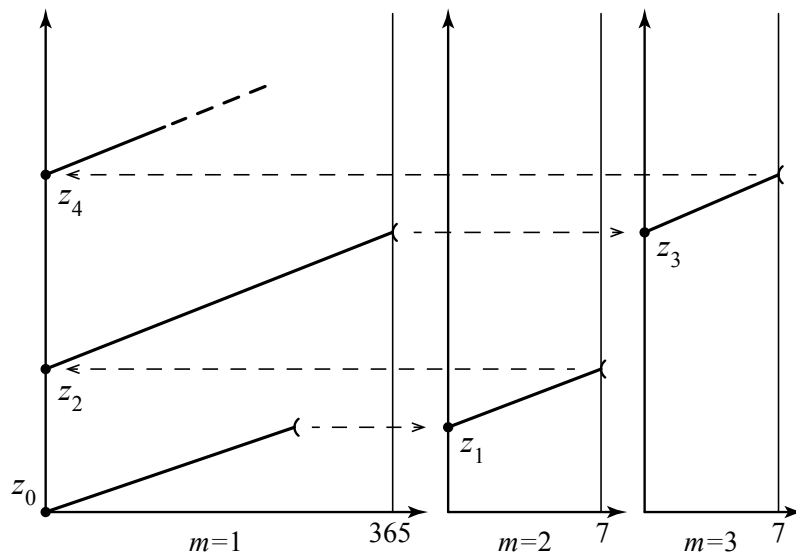


Figure 1. An example trajectory. The process starts from the point Z_0 in mode $m = 1$ (machine in service). The machine may be sent to the workshop for repairs ($m = 2$) or for maintenance ($m = 3$).

Our aim is to find the value of the setting x that maximizes the expected total benefits $B(x)$, that is, the discounted value (for an interest rate ρ) of production minus maintenance and reparation costs over a period $t_f = 5$ years:

$$B^* = \sup_{x \in [0;1]} B(x),$$

where

$$B(x) = E_{(1,0,0)} \left[\int_0^{t_f} e^{-\rho t} (r(x)\mathbb{1}_{\{m_t=1\}} - p(x)\mathbb{1}_{\{m_t=2\}} - q(x)\mathbb{1}_{\{m_t=3\}}) dt \right].$$

We will use the following values $r(x) = x$, $p(x) = 100x^2$, $q(x) = 5$, $s = 7$ days, $\rho = \frac{0.03}{365}$ and λ represents a Weibull distribution with parameters $\alpha = 2$ et $\beta = 600$.

Our assumptions clearly hold so that we may run our numerical method. We first need to find $N \in \mathbb{N}$ such that $\mathbf{P}_{(1,0,0)}(T_N < t_f)$ be small. Monte Carlo simulations lead to the value $N = 18$. For a fixed $x \in [0; 1]$, we will therefore compute $\tilde{J}_N(\tilde{l}, 0)(1, 0, 0)$ where $\tilde{l}(m, \zeta, t) = e^{-\rho t} (r(x)\mathbb{1}_{\{m=1\}} - p(x)\mathbb{1}_{\{m=2\}} - q(x)\mathbb{1}_{\{m=3\}})\mathbb{1}_{\{t \leq t_f\}}$. Finally, notice that we could have chosen r , p and q slightly more generally by allowing them to be time-dependent.

It is important to stress the fact that, once the Markov chain associated to the process is quantized, we will be able to compute the approximation of $B(x)$ almost instantly for any $x \in [0; 1]$ because the same grids are used for every computation. Thanks to this flexibility, we are able to draw the function $x \rightarrow B(x)$ and, thus, to solve the above optimization problem very easily. This is a very important advantage of our method. Indeed, if we computed $B(x)$ through standard methods such as Monte Carlo simulations, we would have to repeat the whole algorithm again and again for each value of x and solving the optimization problem would be intractable.

The following figure represents the approximation of the function B computed on a constant step grid of $[0; 1]$ with step 10^{-2} . This leads to the solution of the earlier optimization problem. Indeed, we obtain $B^* = B(x^*) = 537.84$ where $x^* = 0.78$ is the value of the setting x that maximizes the benefits of the factory.

Now let $x = 0.78$. The following table presents the values of \hat{V}_N , which are the approximations of $B(x)$, for different number of points in the quantization grids. A reference value $B_{\text{Monte Carlo}} = 537.69$ is obtained via the Monte Carlo method (10^8 simulations).

Points in the quantization grids	\hat{V}_N	relative error to 537.69
20 points	542.14	0.83%
50 points	539.57	0.35%
100 points	538.24	0.10%
500 points	537.84	0.03%

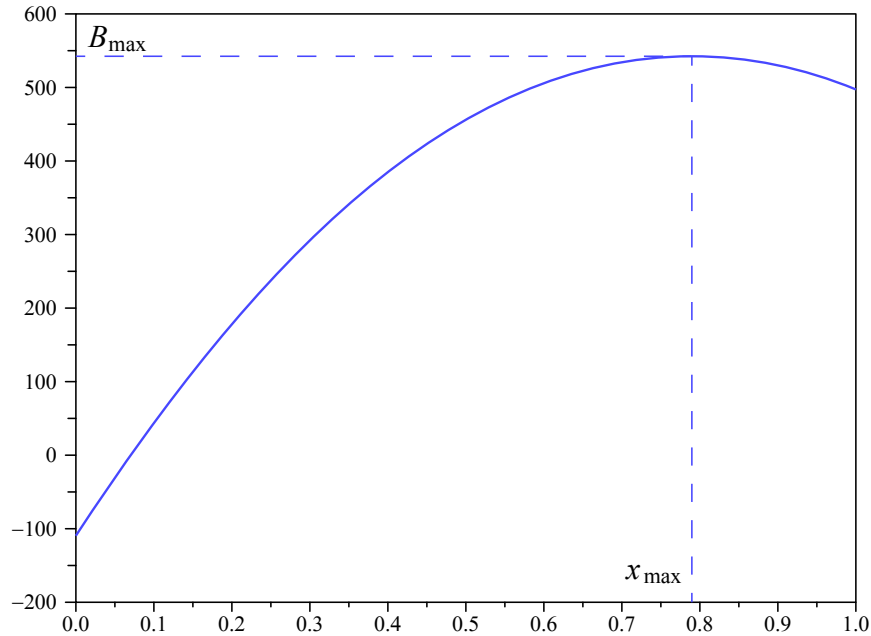


Figure 2. The function B drawn with 500 points in the quantization grids.

From a computational time point of view, we have already explained that the computation of large quantization grids is, by far, the most costly step since it may take up to several hours whereas the approximation of the expectation that follows is then almost instantaneous. However, we may notice, in the above table, that grids containing only 50 points yield a quite accurate result with merely 0.35% error. Such grids only require a few minutes to be designed.

Remark 6.1. We already noticed that the same grids may serve several purposes. For instance, we may also have been interested in the computation of the mean time spent by the machine in the workshop by taking $l(m, \zeta, t) = \mathbb{1}_{\{m \in \{2;3\}\}}$.

6.2. A corrosion model. We consider here a corrosion model for an aluminum metallic structure. This example was provided by Astrium. It concerns a small structure within a strategic ballistic missile. The missile is stored successively in three different environments which are more or less corrosive. It is made to have potentially large storage durations. The requirement for security is very strong. The mechanical stress exerted on the structure depends in part on its thickness. A loss of thickness will cause an overconstraint and therefore increase the risk of rupture. It is thus crucial to study the evolution of the thickness of the structure over time.

Let us describe more precisely the usage profile of the missile. It is stored successively in three different environments: the workshop ($m = 1$), the submarine in operation ($m = 2$) and the submarine in dry-dock ($m = 3$). This is because the

structure must be equipped and used in a given order. Then it goes back to the workshop and so on. The missile stays in each environment during a random duration with exponential distribution. Its parameter λ_m depends on the environment. The degradation law for the thickness loss then depends on the environment through two parameters, a deterministic transition period η_m and a random corrosion rate ρ uniformly distributed within a given range. Typically, the workshop and dry-dock are the most corrosive environments but the time spent in operation is more important. The randomness of the corrosion rate accounts for small variations and uncertainties in the corrosiveness of each environment.

In each environment $m \in \{1; 2; 3\}$, the thickness loss d_m evolves in time as

$$d_m(\rho, s) = \rho(s + \eta_m(e^{-s/(2\eta_m)} - 1)). \quad (10)$$

Here are the numerical values of the parameters of the corrosion model:

	environment 1	environment 2	environment 3
λ_m (h ⁻¹)	(17520) ⁻¹	(131400) ⁻¹	(8760) ⁻¹
η_m (h)	30000	200000	40000
ρ (mm/h)	[10 ⁻⁶ , 10 ⁻⁵]	[10 ⁻⁷ , 10 ⁻⁶]	[10 ⁻⁶ , 10 ⁻⁵]

Initially, the structure is in environment $m = 1$ and the thickness loss is null. One draws the corrosion rate ρ_0 uniformly distributed in the interval [10⁻⁶, 10⁻⁵] and the time of the first change of environment T_1 exponentially distributed with parameter $\lambda_1 = (17520)^{-1}$ hours⁻¹. The corrosion starts according to (10) so that, for all $0 \leq t \leq T_1$, the loss of thickness is $d_1(\rho_0, t)$. The structure then moves to environment 2 and the process restarts similarly: a new corrosion rate ρ_{T_1} is drawn according to a uniform law on [10⁻⁷, 10⁻⁶], the time of the second jump T_2 is drawn so that $T_2 - T_1$ is exponentially distributed with parameter $\lambda_2 = (131400)^{-1}$ hours⁻¹ and for $T_1 \leq t \leq T_2$, the loss of thickness is $d_1(\rho_0, T_1) + d_2(\rho_{T_1}, t - T_1)$ and so on.

At each change of environment, a new corrosion rate ρ is drawn according to a uniform law on the corresponding interval. The thickness loss, however, evolves continuously.

We are interested in computing the mean loss of thickness in environment 2 until a given time $t_f = 18$ years.

Modeling by PDMP.

The state space E. The loss of thickness will be modeled by a PDMP whose modes are the different environments. Let then $M = \{1, 2, 3\}$. The PDMP $(X_t)_{t \geq 0}$ will contain the following components: the mode $m \in M$, the loss of thickness d , the time since the last jump s (this is to ensure that the Markov property is satisfied),

the corrosion rate ρ and the time t (since we consider the time-augmented process). Clearly, one has always $s \leq t$, so we can reasonably consider the state space

$$E = \{(m, d, s, \rho, t) \in M \times \mathbb{R}^+ \times \mathbb{R}^+ \times [10^{-7}; 10^{-5}] \times \mathbb{R}^+ \text{ such that } s \leq t\}.$$

The flow Φ . The flow is given for all $u \geq 0$ by

$$\Phi\left(\begin{pmatrix} m \\ d \\ s \\ \rho \\ t \end{pmatrix}, u\right) = \begin{pmatrix} m \\ d + d_m(\rho, s + u) - d_m(\rho, s) \\ s + u \\ \rho \\ t + u \end{pmatrix}.$$

The transition kernel Q . Let us now study the jumps of this process. When the process jumps from a point $x = (m, d, s, \rho, t) \in E$, m becomes $m + 1$ modulo 3 (denoted $m + 1[3]$), d and t remain unchanged, s becomes 0. Only ρ is randomly drawn, according to a uniform law on an interval $[\rho_{\min}; \rho_{\max}]$ that depends on the new mode. One has then for $w \in B(E)$, $x = (m, d, s, \rho, t) \in E$, and $u \geq 0$,

$$\begin{aligned} Qw\left(\begin{pmatrix} m \\ d \\ s \\ \rho \\ t \end{pmatrix}, u\right) &= Qw\left(\begin{pmatrix} cm \\ d + d_m(\rho, s + u) - d_m(\rho, s) \\ s + u \\ \rho \\ t + u \end{pmatrix}\right) \\ &= \frac{1}{\rho_{\max} - \rho_{\min}} \int_{\rho_{\min}}^{\rho_{\max}} w\left(\begin{pmatrix} m + 1[3] \\ d + d_m(\rho, s + u) - d_m(\rho, s) \\ 0 \\ \tilde{\rho} \\ t + u \end{pmatrix}\right) d\tilde{\rho}. \end{aligned} \quad (11)$$

The cost function l . The function $l \in B(E)$ will be the cost function to compute the mean loss of thickness in mode 2. It is defined as follows: for all $x = (m, d, s, \rho, t) \in E$ and $u \geq 0$,

$$l(\Phi(x, u)) = \rho\left(1 - \frac{1}{2}e^{-(s+u)/(2\eta m)}\right)\mathbb{1}_{\{m=2\}} = \frac{d}{du}(d_m(\rho, s + u))\mathbb{1}_{\{m=2\}}. \quad (12)$$

One then defines $\tilde{l}(\Phi(x, u)) = l(\Phi(x, u))\mathbb{1}_{\{t+u \leq t_f\}}$, so that

$$\begin{aligned} L(x, u) &= \int_0^u \tilde{l}(\Phi(x, u')) du' = \int_0^{u \wedge (t_f - t)^+} l(\Phi(x, u')) du' \\ &= (d_m(\rho, s + u \wedge (t_f - t)^+) - d_m(\rho, s))\mathbb{1}_{\{m=2\}}; \end{aligned}$$

that is indeed the thickness lost in mode 2 from the point $x = (m, d, s, \rho, t)$ during a time $u \wedge (t_f - t)^+$.

The assumptions. Assumptions 2.1 and 2.9 are clearly satisfied. It is easy to check, from (12), that $l \in \mathbf{L}_c(E)$, so Assumption 3.1 holds.

We now turn to Assumption 2.13 and we will see that, although it does not hold for any function $w \in \mathbf{L}_c^v(E)$, it holds for a sufficiently big subclass of functions. We first need to make a remark. Recall that for all $x = (m, d, s, \rho, t) \in E$ and for all $k \in \{0, \dots, N\}$, one has $v_{N-k}(x) = \mathbf{E}_x \text{bigl} l[\int_0^{T_k} l(\Phi(x, u)) \mathbb{1}_{\{t+u \leq t_f\}} du]$. Therefore, for all $k \in \{0, \dots, N\}$ the function v_k as well as the function \tilde{l} satisfy

$$x = (m, d, s, \rho, t) \in E \text{ and } t \geq t_f \implies w(x) = 0. \quad (13)$$

The next step consists in proving that Assumption 2.13, although it is not satisfied for any function $w \in \mathbf{L}_c^v(E)$, holds for any function $w \in \mathbf{L}_c^v(E)$ that also satisfies condition (13). This is done in Lemma 6.2 and it is sufficient because in the proof of the theorem that ensures the convergence of our approximation scheme, Assumption 2.13 is only used with the functions $(v_k)_{k \in \{0, \dots, N\}}$ that do satisfy condition (13).

Lemma 6.2. *There exists $[Q] \in \mathbb{R}^+$ such that for all $v \geq 0$ and $w \in \mathbf{L}_c^v(E)$ that satisfies condition (13), one has for all $x, x' \in E$ and $0 \leq u \leq v$,*

$$|Qw(\Phi(x, u)) - Qw(\Phi(x', u))| \leq [Q][w]_1^{E,v} |x - x'|.$$

Proof. Let $x = (m, d, s, \rho, t)$ and $x' = (m', d', s', \rho', t') \in E$ with for instance $t \leq t'$. First we may choose $m = m'$; otherwise, $|x - x'| = +\infty$ and there is nothing to prove. Now, we are facing three different cases:

- If $t_f \leq t + u \leq t' + u$, then one has $Qw(\Phi(x, u)) = Qw(\Phi(x', u)) = 0$ because w satisfies condition (13) and there is nothing to prove.
- If $t + u \leq t_f \leq t' + u$, notice that

$$Qw(\Phi(x', u)) = Qw(\Phi((m', d', s', \rho', t_f), u)) = 0$$

(this stems from condition (13)), so that we are reduced to the following case.

- We assume from now on that $t + u \leq t' + u \leq t_f$. We now intend to bound $|Qw(\Phi(x, u)) - Qw(\Phi(x', u))|$. It is clear from (11) that we only need to prove that the function $(\rho, s) \rightarrow d_m(\rho, s)$, defined by (10), is Lipschitz continuous with respect to both its variables on the set $[10^{-7}; 10^{-5}] \times [0; t_f]$. Indeed, we have $s \leq t$ and $s' \leq t'$ so that $s, s', s + u, s' + u \leq t_f$. Standard computations yield

$$|d_m(\rho, s) - d_m(\rho', s')| \leq s|\rho - \rho'| + \frac{3}{2}\rho'|s - s'| \leq t_f|\rho - \rho'| + \frac{3}{2}10^{-5}|s - s'|.$$

Hence the result. \square

Assumption 2.10 is not satisfied because in our corrosion model, one has $t^*(x) = +\infty$ for all $x \in E$. Besides, we may notice that the previous proof would have been more straightforward if t^* had been bounded. Indeed in that case, we would have had $s, s', s+u, s'+u \leq C_{t^*}$ and the introduction of condition (13) would have been unnecessary. Nevertheless, we have been able to overcome the drawback of having t^* unbounded by noticing that somehow the deterministic time horizon t_f plays the part of the missing C_{t^*} . This is the meaning of condition (13): roughly speaking, we do not consider what happens beyond t_f .

More generally, we will now see that in our deterministic time horizon problem, the boundedness of t^* may be dropped and our results remain true replacing C_{t^*} by t_f . This is clear in the case of Proposition A.2 because the function \tilde{l} satisfies the condition (13). Proposition A.7 remains also true replacing C_{t^*} by t_f . Indeed, on the one hand, it is clear that $L(x, u) \leq t_f C_l$. On the other hand, when computing $|v_n(\Phi(x, u)) - v_n(\Phi(x', u'))|$, we are facing three cases, as in the proof of Lemma 6.2:

- If $t_f \leq u \leq u'$, one has

$$v_n(\Phi(x, u)) = v_n(\Phi(x', u')) = 0,$$

by condition (13).

- If $u \leq t_f \leq u'$, one has

$$|v_n(\Phi(x, u)) - v_n(\Phi(x', u'))| = |v_n(\Phi(x, u)) - v_n(\Phi(x', t_f))|,$$

since $v_n(\Phi(x', u')) = v_n(\Phi(x', t_f)) = 0$ (condition (13) once again), so that we are reduced to the next case.

- If $u \leq u' \leq t_f$, the computations remain unchanged and t_f replaces C_{t^*} as a bound for u and u' .

Numerical results. The table below presents the values of the loss of thickness in environment 2 obtained through our approximation scheme with quantization grids of varying fineness, as well as the relative deviation with respect to the Monte Carlo value of 0.036755, obtained with 10^8 simulations.

Quantization grids	\hat{V}_0	error	Quantization grids	\hat{V}_0	error
20 points	0.038386	4.43%	2000 points	0.037041	0.77%
50 points	0.037804	2.85%	4000 points	0.037007	0.69%
100 points	0.037525	2.09%	6000 points	0.036973	0.57%
200 points	0.037421	1.81%	8000 points	0.036944	0.49%
500 points	0.037264	1.38%	10000 points	0.036911	0.40%
1000 points	0.037160	1.10%	12000 points	0.036897	0.36%

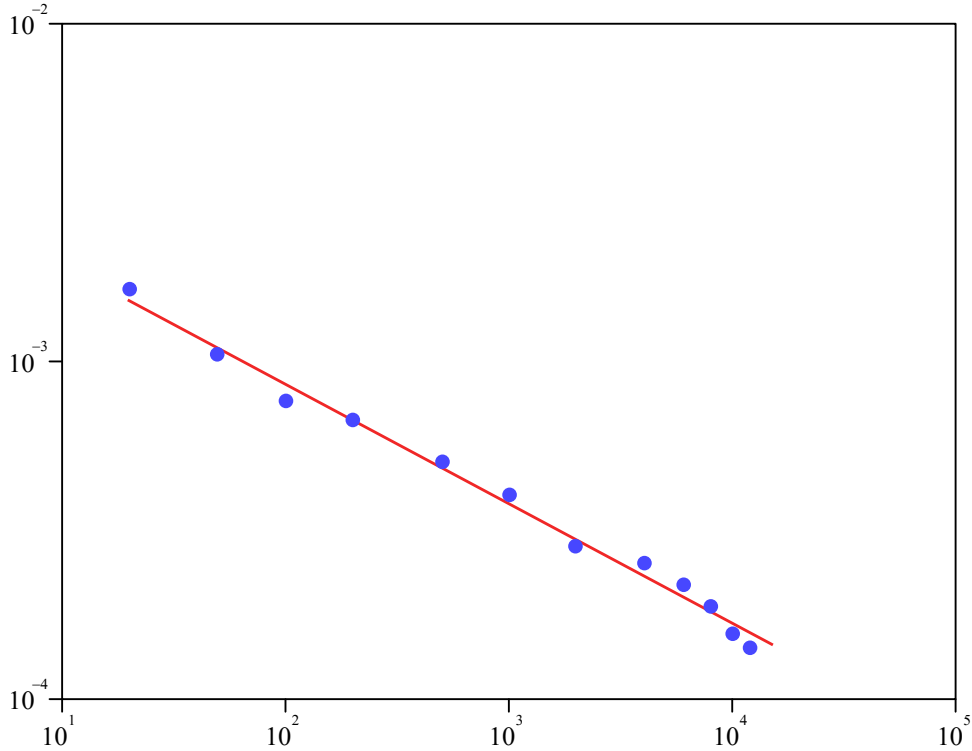


Figure 3. Log-log plot of error when approximating the loss of thickness in environment 2 versus number of points in the quantization grids. The empirical convergence rate, estimated through a regression model, is -0.35 .

Figure 3 presents respectively the empirical convergence rate. The convergence rate, estimated through a regression model is -0.35 . This is roughly the same order of magnitude as the rate of convergence of the optimal quantizer (see for instance [9]) since here the dimension is 3 (indeed, m is deterministic and $s = 0$ immediately after a jump so that we only quantize the variables ρ , d and t).

Finally, we show here the CPU time to compute the expectations from the quantization grids (computations are run with Matlab R2010b on a MacBook Pro 2.66 GHz i7 processor). The CPU time for 10^8 Monte Carlo simulations was approximately 16 000 s. It can be seen that, once the quantization grids are obtained, our approximation scheme performs very fast.

Quantization grids	CPU time (s)	Quantization grids	CPU time (s)
20 points	0.0059	2000 points	1.5
50 points	0.0085	4000 points	5.6
100 points	0.014	6000 points	13
200 points	0.034	8000 points	24
500 points	0.12	10000 points	35
1000 points	0.37	12000 points	54

7. Conclusion

We have presented an efficient and easy to implement numerical method to approximate expectations of functionals of piecewise deterministic Markov processes. We proved the convergence of our algorithm with bounds for the rate of convergence.

Although our method concerns time invariant functionals, we proved that we are able to tackle time-dependent problems such as Lipschitz continuous time-dependent functionals or deterministic time horizon expectations. Indeed, we proved that, thanks to the introduction of the time-augmented process, time-dependent problems may be seen, paradoxically, as special cases of the time invariant situation.

Our method is easy to implement because it merely requires to be able to simulate the process. Furthermore, although the computation of the quantization grids may be quite time-consuming, it may be performed preliminarily because the grids only depend on the dynamics of the process and not on the cost functions l and c . Therefore, they may be stored off-line and serve several purposes. As illustrated by the examples presented in Section 6, storing the grids provides to our approximation scheme efficiency and flexibility. Indeed, the computation of the expectation can be performed very quickly once the grids are available. Thus, if one decides for instance to modify the functional, the same grids may be used so that the new result is obtained very quickly. This flexibility is an important advantage over standard Monte Carlo simulations.

Appendix A. Lipschitz continuity of F , G and v_n

The first lemma and the first proposition of this section present mainly the Lipschitz continuity of the functions δ^A and F . They are stated without proof because they are quite straightforward.

Lemma A.1. *The function δ^A is Lipschitz continuous with respect to both its variables; i.e., for all $x, y \in E$ and $u, t \in \mathbb{R}$, one has*

$$\begin{aligned} |\delta^A(x, t) - \delta^A(y, t)| &\leq A[t^*]|x - y|, \\ |\delta^A(x, t) - \delta^A(x, u)| &\leq A|t - u|, \end{aligned}$$

Moreover, one has for all $x \in E$ and $t, s \geq 0$ such that $t + s \leq t^*(x)$,

$$\delta^A(\Phi(x, s), t) = \delta^A(x, t + s).$$

Proposition A.2. *The function F introduced in Definition 3.3, is Lipschitz continuous with respect to both its variables. For all $x, y \in E$ and $u, v \in [0; t^*(x) \wedge t^*(y)]$, one has*

$$|F(x, u) - F(y, v)| \leq [F]_1|x - y| + [F]_2|u - v|,$$

with

$$[F]_1 = C_{t^*}[l]_1 + [c]_* + A[t^*]C_c, \quad [F]_2 = C_l + AC_c.$$

The next two lemmas are adapted from [6], the second one being a special case of Lemma A.1 there. Thus, they are stated without proof.

Lemma A.3. For $h \in L_c(E)$, $(x, y) \in E^2$, and $t \leq t^*(x) \wedge t^*(y)$

$$\left| \int_t^{t^*(x)} h(\Phi(x, s))e^{-\Lambda(x, s)} ds - \int_t^{t^*(y)} h(\Phi(y, s))e^{-\Lambda(y, s)} ds \right| \leq (C_{t^*}[h]_1 + (C_{t^*}^2[\lambda]_1 + [t^*]C_h)|x - y|.$$

Lemma A.4. For $h \in L_c(\partial E) \cup L_c(E)$ and $x, y \in E$, one has

$$\left| e^{-\Lambda(x, t^*(x))} h(\Phi(x, t^*(x))) - e^{-\Lambda(y, t^*(y))} h(\Phi(y, t^*(y))) \right| \leq ([h]_* + C_h(C_{t^*}[\lambda]_1 + [t^*]C_\lambda))|x - y|.$$

The following notation will be convenient later on. For $w \in L_c(E)$, $x \in E$ and $t \in [0; t^*(x)]$, we define

$$\begin{aligned} G_t w(x) &= E_x[(F(x, S_1) + w(Z_1))\mathbb{1}_{\{S_1 \geq t\}}] \\ &= E_x[(L(x, S_1) + C(x, S_1) + w(Z_1))\mathbb{1}_{\{S_1 \geq t\}}]. \end{aligned}$$

In particular, $G_0 = G$. Since we know the law of (Z_1, S_1) , it can be shown that

$$G_t w(x) = \Upsilon_1(x) + \Upsilon_2(x) + \Upsilon_3(x) + \Upsilon_4(x) + \Upsilon_5(x), \quad (14)$$

with

$$\begin{aligned} \Upsilon_1(x) &= e^{-\Lambda(x, t)} \int_0^t l \circ \Phi(x, s) ds, \\ \Upsilon_2(x) &= \int_t^{t^*(x)} l \circ \Phi(x, s) e^{-\Lambda(x, s)} ds, \\ \Upsilon_3(x) &= c \circ \Phi(x, t^*(x)) \int_t^{t^*(x)} \delta^A(x, s) \lambda \circ \Phi(x, s) e^{-\Lambda(x, s)} ds, \\ \Upsilon_4(x) &= \int_t^{t^*(x)} (\lambda Q w) \circ \Phi(x, s) e^{-\Lambda(x, s)} ds, \\ \Upsilon_5(x) &= e^{-\Lambda(x, t^*(x))} (Q w + c) \circ \Phi(x, t^*(x)). \end{aligned}$$

Proposition A.5. For $w \in L_c(E)$, $(x, y) \in E^2$ and $t \in [0; t^*(x) \wedge t^*(y)]$, one has

$$|G_t w(x) - G_t w(y)| \leq K(A, w)|x - y|,$$

where $K(A, w) = E_1 + E_2 A + E_3[w]_1 + E_4 C_w + [Q][w]_*$, with

$$\begin{aligned} E_1 &= 2[l]_1 C_{t^*} + C_l([t^*] + 2C_{t^*}^2[\lambda]_1) + [c]_*(1 + C_{t^*} C_\lambda) \\ &\quad + C_c(2[\lambda]_1 C_{t^*} + C_\lambda C_{t^*}^2[\lambda]_1 + 2[t^*] C_\lambda), \\ E_2 &= C_c C_{t^*} C_\lambda [t^*], \\ E_3 &= (1 + C_{t^*} C_\lambda)[Q], \\ E_4 &= 2C_\lambda [t^*] + C_{t^*}[\lambda]_1(2 + C_{t^*} C_\lambda). \end{aligned}$$

Proof. Let $w \in L_c(E)$, $(x, y) \in E^2$ and $t \in [0; t^*(x) \wedge t^*(y)]$. In view of (14), we naturally split $|G_t w(x) - G_t w(y)|$ into the sum of five differences.

The first one, $|\Upsilon_1(x) - \Upsilon_1(y)|$, is bounded by

$$\begin{aligned} |\Upsilon_1(x) - \Upsilon_1(y)| &\leq C_{t^*} C_l |e^{-\Lambda(x,t)} - e^{-\Lambda(y,t)}| + \int_0^t (l \circ \Phi(x, s) - l \circ \Phi(y, s)) ds \\ &\leq (C_{t^*}^2 C_l [\lambda]_1 + C_{t^*} [l]_1) |x - y|. \end{aligned}$$

The differences $|\Upsilon_2(x) - \Upsilon_2(y)|$ and $|\Upsilon_4(x) - \Upsilon_4(y)|$ can be bounded thanks to Lemma A.3, with successively $h = l$ and $h = \lambda Qw$. Notice that $C_\lambda Qw \leq C_\lambda C_w$ and $[\lambda Qw]_1 \leq C_\lambda [Q][w]_1 + C_w [\lambda]_1$.

For the difference of the Υ_5 terms, we use Lemma A.4 with $h = Qw + c$. Notice that $C_{Qw+c} \leq C_w + C_c$ and that $[Qw + c]_* \leq [Q]([w]_* + [w]_1) + [c]_*$.

Finally, to bound $|\Upsilon_3(x) - \Upsilon_3(y)|$, we assume without loss of generality that $t^*(x) \leq t^*(y)$ and we have

$$\begin{aligned} &|\Upsilon_3(x) - \Upsilon_3(y)| \\ &\leq C_c \int_t^{t^*(x)} |\delta^A(x, s) \lambda \circ \Phi(x, s) e^{-\Lambda(x,s)} - \delta^A(y, s) \lambda \circ \Phi(y, s) e^{-\Lambda(y,s)}| ds \\ &\quad + C_c \int_{t^*(x)}^{t^*(y)} |\delta^A(y, s) \lambda \circ \Phi(y, s) e^{-\Lambda(y,s)}| ds + [c]_* C_{t^*} C_\lambda |x - y| \\ &\leq C_c \int_t^{t^*(x)} (C_\lambda |\delta^A(x, s) - \delta^A(y, s)| + [\lambda]_1 |x - y| + C_\lambda |e^{-\Lambda(x,s)} - e^{-\Lambda(y,s)}|) ds \\ &\quad + C_c [t^*] C_\lambda |x - y| + [c]_* C_{t^*} C_\lambda |x - y| \\ &\leq (C_c C_{t^*} (C_\lambda A[t^*] + [\lambda]_1 + C_\lambda C_{t^*} [\lambda]_1) + C_c [t^*] C_\lambda + [c]_* C_{t^*} C_\lambda) |x - y|. \end{aligned}$$

The result follows. \square

The next lemma is stated without proof, as it is very close to [5, Lemma 51.7].

Lemma A.6. *For all $x \in E$ and $t \in [0; t^*(x)]$, one has*

$$v_n(\Phi(x, t)) = e^{\Lambda(x,t)} G_t v_{n-1}(x) - \int_0^t l \circ \Phi(x, s) ds.$$

Proposition A.7. *For all $n \in \{0, 1, \dots, N\}$, one has $v_n \in \mathbf{L}_c(E)$ and*

$$\begin{aligned} C_{v_n} &\leq n(C_{t^*}C_l + C_c), \\ [v_n]_1 &\leq e^{C_{t^*}C_\lambda} (K(A, v_{n-1}) + nC_{t^*}[\lambda]_1(C_{t^*}C_l + C_c)) + C_{t^*}[l]_1, \\ [v_n]_2 &\leq e^{C_{t^*}C_\lambda} (C_{t^*}C_lC_\lambda + 2C_l + C_\lambda C_c + (2n-1)C_\lambda(C_{t^*}C_l + C_c)) + C_l, \\ [v_n]_* &\leq [v_n]_1 + [t^*][v_n]_2, \\ [v_n] &\leq K(A, v_{n-1}), \end{aligned}$$

Proof. Recall that for $x \in E$, one has from Definition 3.3

$$v_n(x) = Gv_{n-1}(x) = \mathbf{E}_x [L(x, S_1)] + \mathbf{E}_x [C(x, S_1)] + \mathbf{E}_x [v_{n-1}(Z_1)].$$

Thus, $C_{v_n} \leq C_{t^*}C_l + C_c + C_{v_{n-1}} \leq n(C_{t^*}C_l + C_c)$ by induction.

Let us now turn to $[v_n]_1$. Lemma A.6 yields

$$\begin{aligned} &|v_n(\Phi(x, t)) - v_n(\Phi(y, t))| \\ &\leq |e^{\Lambda(x, t)} G_t v_{n-1}(x) - e^{\Lambda(y, t)} G_t v_{n-1}(y)| + \int_0^t |l \circ \Phi(x, s) - l \circ \Phi(y, s)| ds \\ &\leq e^{\Lambda(x, t)} |G_t v_{n-1}(x) - G_t v_{n-1}(y)| + |G_t v_{n-1}(y)| |e^{\Lambda(x, t)} - e^{\Lambda(y, t)}| \\ &\hspace{15em} + C_{t^*}[l]_1 |x - y|. \end{aligned}$$

The result follows using Proposition A.5 and noticing that

$$\begin{aligned} \Lambda(x, t) &\leq C_{t^*}C_\lambda, \\ |G_t v_{n-1}(y)| &\leq C_{t^*}C_l + C_c + C_{v_{n-1}} \leq n(C_{t^*}C_l + C_c), \\ |e^{\Lambda(x, t)} - e^{\Lambda(y, t)}| &\leq e^{C_{t^*}C_\lambda} C_{t^*}[\lambda]_1 |x - y|. \end{aligned}$$

We now turn to $[v_n]_2$. For $x \in E$ and $s, t \in [0, t^*(x)]$ with $s \leq t$, one has

$$\begin{aligned} |v_n(\Phi(x, t)) - v_n(\Phi(x, s))| &\leq e^{\Lambda(x, t)} |G_t v_{n-1}(x) - G_s v_{n-1}(x)| \\ &\quad + |G_s v_{n-1}(x)| |e^{\Lambda(x, t)} - e^{\Lambda(x, s)}| + C_l |t - s|. \end{aligned}$$

Moreover, from (14), one has

$$\begin{aligned}
 & |G_t v_{n-1}(x) - G_s v_{n-1}(x)| \\
 & \leq E_x [|F(x, S_1) + v_{n-1}(Z_1)| \mathbb{1}_{\{s \leq S_1 < t\}}] \\
 & \leq \left| e^{-\Lambda(x,t)} \int_0^t l(\Phi(x, u)) du - e^{-\Lambda(x,s)} \int_0^s l(\Phi(x, u)) du \right| \\
 & \quad + \int_s^t |l(\Phi(x, u)) e^{-\Lambda(x,u)}| du \\
 & \quad + |c \circ \Phi(x, t^*(x))| \int_s^t |\delta^A(x, u) \lambda \circ \Phi(x, u) e^{-\Lambda(x,u)}| du \\
 & \quad + \int_s^t |(\lambda Q v_{n-1}) \circ \Phi(x, u) e^{-\Lambda(x,u)}| du \\
 & \leq (C_{t^*} C_l |e^{-\Lambda(x,t)} - e^{-\Lambda(x,s)}| + C_l |t - s|) + (C_l |t - s|) \\
 & \quad + (C_c C_\lambda |t - s|) + (C_\lambda C_{v_{n-1}} |t - s|)
 \end{aligned}$$

and

$$|e^{\Lambda(x,t)} - e^{\Lambda(x,s)}| \leq e^{C_{t^*} C_\lambda} C_\lambda |t - s|.$$

Finally, the bound for $[v_n]$ is a direct consequence from Proposition A.5. \square

Appendix B. Relaxed assumption on the running cost function

In this section, we consider the approximation applied to the time-augmented process so that the local characteristics are $\tilde{\Phi}$, $\tilde{\lambda}$ and \tilde{Q} defined in Section 5.1. Moreover, we consider a function $l \in \mathbf{L}_c(\tilde{E})$ and we define $\tilde{l} \in B(\tilde{E})$ by

$$\text{for all } \xi = (x, t) \in \tilde{E}, \tilde{l}(\xi) = l(x, t) \mathbb{1}_{\{t \leq t_f\}}.$$

We intend to prove that the convergence of our approximation scheme, stated by Theorem 4.5, remains true if we choose \tilde{l} as the running cost function even though it does not fulfill the required Lipschitz conditions, i.e., $\tilde{l} \notin \mathbf{L}_c(\tilde{E})$. Indeed, the Lipschitz continuity of l is used four times in the proof of the theorem, once in Proposition A.2, twice in Proposition A.5 (when bounding the difference of the Υ_1 terms and the one of the Υ_2 ones) and once in Proposition A.7 (when bounding $[v_n]_1$). In each case, the Lipschitz continuity of the running cost function l is used to bound a term of the form

$$\int_s^{s'} |\tilde{l} \circ \tilde{\Phi}(\xi, u) - \tilde{l} \circ \tilde{\Phi}(\xi', u)| du \quad (15)$$

for $\xi, \xi' \in \tilde{E}$ and $s, s' \in [0; \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')]$, or of the form

$$\int_s^{\tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')} |\tilde{l} \circ \tilde{\Phi}(\xi, u) e^{-\tilde{\Lambda}(\xi, u)} - \tilde{l} \circ \tilde{\Phi}(\xi', u) e^{-\tilde{\Lambda}(\xi', u)}| du \quad (16)$$

for $\xi, \xi' \in \tilde{E}$ and $s \in [0; \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')]$ and where we use the natural notation $\tilde{\Lambda}(\xi, u) = \int_0^u \tilde{\lambda}(\tilde{\Phi}(\xi, v)) dv$. Concerning this second form, Equation (16), notice that

$$\begin{aligned} & \int_s^{\tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')} |\tilde{l} \circ \tilde{\Phi}(\xi, u) e^{-\tilde{\Lambda}(\xi, u)} - \tilde{l} \circ \tilde{\Phi}(\xi', u) e^{-\tilde{\Lambda}(\xi', u)}| du \\ & \leq \int_s^{\tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')} |\tilde{l} \circ \tilde{\Phi}(\xi, u) - \tilde{l} \circ \tilde{\Phi}(\xi', u)| du \\ & \quad + C_l \int_s^{\tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')} |e^{-\tilde{\Lambda}(\xi, u)} - e^{-\tilde{\Lambda}(\xi', u)}| du \\ & \leq \int_s^{\tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')} |\tilde{l} \circ \tilde{\Phi}(\xi, u) - \tilde{l} \circ \tilde{\Phi}(\xi', u)| du + C_l C_{\tilde{t}^*}^2[\lambda]_1 |\xi - \xi'|, \end{aligned}$$

so that, to ensure that Theorem 4.5 remains true with \tilde{l} as the running cost function, it is sufficient to be able to bound terms of the form (15). This is done in the following lemma.

Lemma B.1. *For $\xi = (x, t), \xi' = (x', t') \in \tilde{E}$ and $s \in [0; \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')]$, one has*

$$\int_0^s |\tilde{l} \circ \tilde{\Phi}(\xi, u) - \tilde{l} \circ \tilde{\Phi}(\xi', u)| du \leq (C_{\tilde{t}^*}[l]_1 + C_l) |\xi - \xi'|.$$

Proof. Let $\xi = (x, t), \xi' = (x', t') \in \tilde{E}$ and $s \in [0; \tilde{t}^*(\xi) \wedge \tilde{t}^*(\xi')]$. One has

$$\begin{aligned} & \int_0^s |\tilde{l} \circ \tilde{\Phi}(\xi, u) - \tilde{l} \circ \tilde{\Phi}(\xi', u)| du \\ & \leq \int_0^s |l \circ \tilde{\Phi}(\xi, u) \mathbb{1}_{\{t+u \leq t_f\}} - l \circ \tilde{\Phi}(\xi', u) \mathbb{1}_{\{t'+u \leq t_f\}}| du \\ & \leq \int_0^s |l \circ \tilde{\Phi}(\xi, u) - l \circ \tilde{\Phi}(\xi', u)| du + C_l \int_0^s |\mathbb{1}_{\{t+u \leq t_f\}} - \mathbb{1}_{\{t'+u \leq t_f\}}| du \end{aligned}$$

The left-hand side term is bounded by $C_{\tilde{t}^*}[l]_1 |\xi - \xi'|$ since $l \in \mathbf{L}_c(\tilde{E})$. For the right-hand side term, assume without loss of generality that $t \leq t'$, one has

$$|\mathbb{1}_{\{t+u \leq t_f\}} - \mathbb{1}_{\{t'+u \leq t_f\}}| = |\mathbb{1}_{\{t-t_f \leq u\}} - \mathbb{1}_{\{t'-t_f \leq u\}}| = \mathbb{1}_{\{t-t_f \leq u < t'-t_f\}},$$

so that the right-hand side term is bounded by $C_l |t - t'| \leq C_l |\xi - \xi'|$. The result follows. \square

Theorem 4.5 remains true if we choose \tilde{l} as the running cost function. One only needs to slightly modify the Lipschitz constants given in propositions A.2, A.5 and A.7. The terms $C_{t^*}[l]_1$ have to be replaced by $C_{t^*}[l]_1 + C_l$.

Appendix C. Proof of Theorem 4.5

The Lipschitz continuity of the functions v_k is proved by Proposition A.7. Now let $A > 0$ and notice that

$$|J_N(l, c)(x) - \hat{V}_0| \leq |J_N(l, c)(x) - V_0| + |V_0 - \hat{V}_0|.$$

Proposition 3.2 says that $|J_N(l, c)(x) - V_0| \leq NC_c C_\lambda / A$ since $V_0 = J_N^A(l, c)(x)$. We now have to bound $|V_0 - \hat{V}_0|$.

Some of the arguments of the proof are similar to the ones used in Theorem 5.1 from [6], thus we will not develop the details of the proof. Recall that $\|V_N - \hat{V}_N\|_p = 0$ and let $k \in \{0, \dots, N-1\}$. In order to bound the approximation error, let us split it into three terms $\|V_k - \hat{V}_k\|_p \leq \Xi_1 + \Xi_2 + \Xi_3$, where

$$\begin{aligned} \Xi_1 &= \|v_k(Z_k) - v_k(\hat{Z}_k)\|_p, \\ \Xi_2 &= \|Gv_{k+1}(\hat{Z}_k) - \hat{G}_{k+1}v_{k+1}(\hat{Z}_k)\|_p, \\ \Xi_3 &= \|\hat{G}_{k+1}v_{k+1}(\hat{Z}_k) - \hat{G}_{k+1}\hat{v}_{k+1}(\hat{Z}_k)\|_p. \end{aligned}$$

The theorem is then a direct consequence from the three following lemmas, stated without proof, that provide bounds for each of these three terms.

Lemma C.1. *The first term, Ξ_1 , is bounded by*

$$\|v_k(Z_k) - v_k(\hat{Z}_k)\|_p \leq [v_k] \|Z_k - \hat{Z}_k\|_p.$$

Lemma C.2. *The second term, Ξ_2 , is bounded by*

$$\begin{aligned} &\|Gv_{k+1}(\hat{Z}_k) - \hat{G}_{k+1}v_{k+1}(\hat{Z}_k)\|_p \\ &\leq [v_{k+1}] \|Z_{k+1} - \hat{Z}_{k+1}\|_p + ([v_k] + [F]_1) \|Z_k - \hat{Z}_k\|_p + [F]_2 \|S_{k+1} - \hat{S}_{k+1}\|_p. \end{aligned}$$

Lemma C.3. *The third term, Ξ_3 , is bounded by*

$$\begin{aligned} &\|\hat{G}_{k+1}v_{k+1}(\hat{Z}_k) - \hat{G}_{k+1}\hat{v}_{k+1}(\hat{Z}_k)\|_p \\ &\leq [v_{k+1}] \|Z_{k+1} - \hat{Z}_{k+1}\|_p + \|V_{k+1} - \hat{V}_{k+1}\|_p. \end{aligned}$$

Acknowledgements

This work was supported by ARPEGE program of the French National Agency of Research (ANR), project "FAUTOCOES", number ANR-09-SEGI-004. Besides, the authors gratefully acknowledge Astrium for its financial support.

References

- [1] V. Bally and G. Pagès, *A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems*, *Bernoulli* **9** (2003), no. 6, 1003–1049. MR 2005f:60096
- [2] V. Bally, G. Pagès, and J. Printems, *A quantization tree method for pricing and hedging multidimensional American options*, *Math. Finance* **15** (2005), no. 1, 119–168. MR 2005k:91142
- [3] A. Brandejsky, B. de Saporta, and F. Dufour, *Numerical methods for the exit time of a piecewise-deterministic markov process*, *Adv. Appl. Probab.* **44** (2012), no. 1. arXiv 1012.2659
- [4] C. Coccozza-Thivent, R. Eymard, and S. Mercier, *A finite-volume scheme for dynamic reliability models*, *IMA J. Numer. Anal.* **26** (2006), no. 3, 446–471. MR 2008h:60370 Zbl 1109.65011
- [5] M. H. A. Davis, *Markov models and optimization*, *Monographs on Statistics and Applied Probability*, no. 49, Chapman & Hall, London, 1993. MR 96b:90002 Zbl 0780.60002
- [6] B. de Saporta, F. Dufour, and K. Gonzalez, *Numerical method for optimal stopping of piecewise deterministic Markov processes*, *Ann. Appl. Probab.* **20** (2010), no. 5, 1607–1637. MR 2011g:60079 Zbl 1197.93162
- [7] R. Eymard, S. Mercier, and A. Prignet, *An implicit finite volume scheme for a scalar hyperbolic problem with measure data related to piecewise deterministic Markov processes*, *J. Comput. Appl. Math.* **222** (2008), no. 2, 293–323. MR 2009j:65194 Zbl 1158.65008
- [8] R. M. Gray and D. L. Neuhoff, *Quantization*, *IEEE Trans. Inform. Theory* **44** (1998), no. 6, 2325–2383. MR 99i:94029 Zbl 1016.94016
- [9] G. Pagès, H. Pham, and J. Printems, *Optimal quantization methods and applications to numerical problems in finance*, *Handbook of computational and numerical methods in finance*, Birkhäuser, Boston, 2004, pp. 253–297. MR 2083055 Zbl 1138.91467

Received May 6, 2011. Revised November 8, 2011.

ADRIEN BRANDEJSKY: `adrien.brandejsky@math.u-bordeaux1.fr`
 INRIA Bordeaux Sud-Ouest, team CQFD, F-33400 Talence, France

and

Université de Bordeaux, IMB, UMR 5251, F-33400 Talence, France

and

CNRS, IMB, UMR 5251, F-33400 Talence, France

BENOÎTE DE SAPORTA: `saporta@math.u-bordeaux1.fr`
 Université de Bordeaux, Gretha, UMR 5113, F-33400 Talence, France

and

CNRS, Gretha, UMR 5113, IMB, UMR 5251, F-33400 Talence, France

and

INRIA Bordeaux Sud-Ouest, team CQFD, F-33400 Talence, France

FRANÇOIS DUFOUR: `francois.dufour@math.u-bordeaux1.fr`
 Université de Bordeaux, IMB, UMR 5251, F-33400 Talence, France

and

CNRS, IMB, UMR 5251, F-33400 Talence, France

and

INRIA Bordeaux Sud-Ouest, team CQFD, F-33400 Talence, France

Communications in Applied Mathematics and Computational Science

map.berkeley.edu/camcos

EDITORS

MANAGING EDITOR

John B. Bell
Lawrence Berkeley National Laboratory, USA
jbbell@lbl.gov

BOARD OF EDITORS

Marsha Berger	New York University berger@cs.nyu.edu	Ahmed Ghoniem	Massachusetts Inst. of Technology, USA ghoniem@mit.edu
Alexandre Chorin	University of California, Berkeley, USA chorin@math.berkeley.edu	Raz Kupferman	The Hebrew University, Israel raz@math.huji.ac.il
Phil Colella	Lawrence Berkeley Nat. Lab., USA pcolella@lbl.gov	Randall J. LeVeque	University of Washington, USA rjl@amath.washington.edu
Peter Constantin	University of Chicago, USA const@cs.uchicago.edu	Mitchell Luskin	University of Minnesota, USA luskin@umn.edu
Maksymilian Dryja	Warsaw University, Poland maksymilian.dryja@acn.waw.pl	Yvon Maday	Université Pierre et Marie Curie, France maday@ann.jussieu.fr
M. Gregory Forest	University of North Carolina, USA forest@amath.unc.edu	James Sethian	University of California, Berkeley, USA sethian@math.berkeley.edu
Leslie Greengard	New York University, USA greengard@cims.nyu.edu	Juan Luis Vázquez	Universidad Autónoma de Madrid, Spain juanluis.vazquez@uam.es
Rupert Klein	Freie Universität Berlin, Germany rupert.klein@pik-potsdam.de	Alfio Quarteroni	Ecole Polytech. Féd. Lausanne, Switzerland alfio.quarteroni@epfl.ch
Nigel Goldenfeld	University of Illinois, USA nigel@uiuc.edu	Eitan Tadmor	University of Maryland, USA etadmor@cscamm.umd.edu
	Denis Talay	INRIA, France denis.talay@inria.fr	

PRODUCTION

contact@msp.org

Silvio Levy, Scientific Editor

Sheila Newbery, Senior Production Editor

See inside back cover or msp.berkeley.edu/camcos for submission instructions.

The subscription price for 2012 is US \$75/year for the electronic version, and \$105/year for print and electronic. Subscriptions, requests for back issues from the last three years and changes of subscribers address should be sent to Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840, USA.

Communications in Applied Mathematics and Computational Science, at Mathematical Sciences Publishers, Department of Mathematics, University of California, Berkeley, CA 94720-3840 is published continuously online. Periodical rate postage paid at Berkeley, CA 94704, and additional mailing offices.

CAMCoS peer review and production are managed by EditFLOW™ from Mathematical Sciences Publishers.

PUBLISHED BY
 **mathematical sciences publishers**
<http://msp.org/>

A NON-PROFIT CORPORATION

Typeset in L^AT_EX

Copyright ©2012 by Mathematical Sciences Publishers

Communications in Applied Mathematics and Computational Science

vol. 7

no. 1

2012

- An embedded boundary method for the Navier–Stokes equations on a time-dependent domain 1
GREGORY H. MILLER and DAVID TREBOTICH
- Slender body theory for Stokes flows with regularized forces 33
RICARDO CORTEZ and MICHAEL NICHOLAS
- Numerical method for expectations of piecewise deterministic Markov processes 63
ADRIEN BRANDEJSKY, BENOÎTE DE SAPORTA and FRANÇOIS DUFOR
- Toward an efficient parallel in time method for partial differential equations 105
MATTHEW EMMETT and MICHAEL L. MINION



1559-3940(2012)7:1;1-4

- 3.2 [BdSD12b] A. BRANDEJSKY, B. DE SAPORTA, AND F. DUFOUR. Numerical methods for the exit time of a piecewise-deterministic markov process. *Advances in Applied Probability*, 44(1) :196–225, 2012.

NUMERICAL METHODS FOR THE EXIT TIME OF A PIECEWISE-DETERMINISTIC MARKOV PROCESS

ADRIEN BRANDEJSKY,* *Université Bordeaux, IMB and INRIA Bordeaux Sud-Ouest*

BENOÎTE DE SAPORTA,* *Université Bordeaux, Gretha and INRIA Bordeaux Sud-Ouest*

FRANÇOIS DUFOUR,** *Université Bordeaux, IMB and INRIA Bordeaux Sud-Ouest*

Abstract

We present a numerical method to compute the survival function and the moments of the exit time for a piecewise-deterministic Markov process (PDMP). Our approach is based on the quantization of an underlying discrete-time Markov chain related to the PDMP. The approximation we propose is easily computable and is even flexible with respect to the exit time we consider. We prove the convergence of the algorithm and obtain bounds for the rate of convergence in the case of the moments. We give an academic example and a model from the reliability field to illustrate the results of the paper.

Keywords: Exit time; piecewise-deterministic Markov process; quantization; numerical method

2010 Mathematics Subject Classification: Primary 60J25; 65C20

Secondary 60K10

1. Introduction

The aim of this paper is to propose a practical numerical method to approximate the survival function and the moments of the exit time for a piecewise-deterministic Markov process based on the quantization of a discrete-time Markov chain naturally embedded within the continuous-time process.

Piecewise-deterministic Markov processes (PDMPs) were introduced by Davis [5] as a general class of stochastic models. PDMPs are a family of Markov processes involving deterministic motion punctuated by random jumps. The motion depends on three local characteristics, namely the flow Φ , the jump rate λ , and the transition measure Q , which specifies the post-jump location. Starting from the point x , the motion of the process follows the flow $\Phi(x, t)$ until the first jump time T_1 , which occurs either spontaneously in a Poisson-like fashion with rate $\lambda(\Phi(x, t))$ or when the flow $\Phi(x, t)$ hits the boundary of the state space. In either case, the location of the process at the jump time T_1 is selected by the transition measure $Q(\cdot, \Phi(x, T_1))$ and the motion restarts from this new point $X(T_1)$ denoted by Z_1 . We similarly define the time S_2 until the next jump; the next jump time is $T_2 = T_1 + S_2$, the next post-jump location $Z_2 = X(T_2)$, and so on. Thus, associated to the PDMP we have discrete-time Markov chains $(Z_n, T_n)_{n \in \mathbb{N}}$, given by the post-jump locations and the jump times, and $(Z_n, S_n)_{n \in \mathbb{N}}$, given by the post-jump locations and the inter-jump times. Suitable choices of the state space and the local characteristics Φ , λ , and Q provide stochastic models covering a great number of problems in operations research; see, for example, [4], [5], and the corrosion model presented in this paper.

Received 6 January 2011; revision received 26 September 2011.

* Postal address: INRIA Bordeaux Sud-Ouest, CQFD Team, 351 cours de la Libération, F-33405 Talence, France.

** Email address: dufour@math.u-bordeaux1.fr

Numerical computation of the moments of the exit time for a Markov process has been studied by Helmes *et al.* [8]. Starting from an assumption related to the generator of the process, they derived a system of linear equations satisfied by the moments. In addition to these equations, they included finitely many Hausdorff moment conditions that are also linear constraints. This optimization problem is a standard linear programming problem for which a lot of efficient software is available. Lasserre and Prieto-Rumeau [9] introduced a similar method, but they improved the efficiency of the algorithm by replacing the Hausdorff moment conditions with semidefinite positivity constraints of some moment matrices. Nevertheless, their approach cannot be applied to PDMPs because the assumption related to the generator of the process is generally not satisfied. In [5, Section 33] Davis gave an iterative method to compute the mean exit time for a PDMP, but his approach involved solving a large set of ordinary differential equations whose forms are very problem specific, depending on the behaviour of the process at the boundary of the state space. Besides, and in the context of applications to reliability, it seems important to also study the distribution of the exit time.

There exists extensive literature on quantization methods for random variables and processes. The interested reader is referred to, e.g. [7], [10], and the references therein. Quantization methods have been developed recently for numerical probability or optimal stochastic control problems with applications in finance (see, e.g. [1]–[3] and [10]). The quantization of a Markov chain $(\Theta_n)_{n \in \mathbb{N}}$ consists in finding, for each n , an optimally designed discretization of the state space of Θ_n providing the best possible L^p -approximation by a random variable $\widehat{\Theta}_n$ taking its values in a grid Γ_n of finite and fixed size as well as a transition measure of the quantized chain $(\widehat{\Theta}_n)_{n \in \mathbb{N}}$. As explained for instance in [10, Section 3], provided that the Markov kernel is Lipschitz, bounds for the rate of L^p -convergence of the quantized process towards the original process are obtained.

In the present work, we consider a PDMP $(X_t)_{t \geq 0}$ with state space E and we present approximation methods to compute the moments and the survival function of the exit time from a set $U \subset E$, given that the PDMP exits the set U before the N th jump time T_N . Roughly speaking, we estimate the moments and the survival function for $\tau \wedge T_N$. In our approach, the first step consists in expressing the j th moment (respectively the survival function) as the last term of some sequence $(p_{k,j})_{k \leq N}$ (respectively $(p_k)_{k \leq N}$) satisfying a recursion $p_{k+1,j} = \psi(p_{k,j})$ (respectively $p_{k+1} = \psi(p_k)$) specifically built within our paper.

In this context, a natural way to deal with these problems is to follow the idea developed in [6], namely to write the recursions in terms of an underlying discrete-time Markov chain and to replace it by its quantized approximation. The definitions of $(p_{k,j})_k$ and $(p_k)_k$ involve some discontinuities related to indicator functions, but, as in [6], we show that they occur with small enough probabilities. However, an important feature that distinguishes the present work from [6] and which prevents a straightforward application of the ideas developed therein, is that an additional important difficulty appears in the definitions of the sequences $(p_{k,j})_k$ and $(p_k)_k$. Indeed, the mapping ψ such that $p_{k+1,j} = \psi(p_{k,j})$ and $p_{k+1} = \psi(p_k)$ is not Lipschitz continuous. One of the main results of this paper is to overcome this difficulty by deriving new and important properties of the Markov chain $(Z_n, T_n)_{n \in \mathbb{N}}$, combined with a sharp feature of the quantization algorithm. We are able to prove the convergence of the approximation scheme. Moreover, in the case of the moments, we even obtain bounds for the rate of convergence. It is important to stress that these assumptions are quite reasonable with regards to the applications.

An important advantage of our method is that it is flexible. Indeed, as pointed out in [1], a quantization-based method is ‘obstacle free’, which means, in our case, that it produces, once and for all, a discretization of the process independently of the set U . Consequently, the

approximation schemes for both the moments and the distribution of the exit time are flexible with respect to U . Indeed, if we are interested in the exit time from a new set U' , it will be possible, provided that U' satisfies the same assumptions as U , to obtain in a very simple way the moments and the distribution of this new exit time. Indeed, the quantization grids are computed only once, stored offline, and may therefore serve many purposes.

The paper is organized as follows. We first recall the definition of a PDMP and state our assumptions. In Section 3, we introduce the moments and the distribution problems, and present recursive methods to solve them. Section 4 contains the main contributions of this paper, namely the approximation schemes, the proofs of convergence, and bounds for the rates of convergence. Two numerical examples are developed in Section 5 and the advantages of our approach are discussed in Section 6.

2. Definitions and assumptions

For any metric space X , we denote by $\mathcal{B}(X)$ its Borel σ -field and by $B(X)$ the set of real-valued, bounded, and measurable functions defined on X . For $a, b \in \mathbb{R}$, $a \wedge b = \min(a, b)$ and $a \vee b = \max(a, b)$.

2.1. Definition of a PDMP

In this section we define a PDMP and introduce some general assumptions. Let M be a finite set, called the set of the modes, that represents the different regimes of evolution of the PDMP (M is supposed to be a finite space although it could be countable); for each $m \in M$, the process evolves in E_m , an open subset of $\mathbb{R}^{d(m)}$ (where $d: M \rightarrow \mathbb{N}^*$). Let

$$E = \{(m, \xi), m \in M, \xi \in E_m\}.$$

This is the state space of the process $(X_t)_{t \in \mathbb{R}^+} = (m_t, \xi_t)_{t \in \mathbb{R}^+}$. Let ∂E be its boundary, let \bar{E} be its closure, and, for any subset Y of E , let Y^c denote its complement.

Define on E the following distance: for $x = (m, \xi)$ and $x' = (m', \xi') \in E$,

$$|x - x'| = \begin{cases} +\infty & \text{if } m \neq m', \\ |\xi - \xi'| & \text{otherwise.} \end{cases}$$

Moreover, for any $x \in E$ and $Y \subset E$, denote by $d(x, Y)$ the distance between the point x and the set Y , i.e. $d(x, Y) = \inf_{y \in Y} |x - y|$.

A PDMP is defined by its local characteristics $(\Phi_m, \lambda_m, Q_m)_{m \in M}$.

- For each $m \in M$, $\Phi_m: \mathbb{R}^{d(m)} \times \mathbb{R} \rightarrow \mathbb{R}^{d(m)}$ is a continuous function called the flow in mode m . For all $t \in \mathbb{R}$, $\Phi_m(\cdot, t)$ is a homeomorphism and $t \rightarrow \Phi_m(\cdot, t)$ is a group, i.e. for all $\xi \in \mathbb{R}^{d(m)}$, $\Phi_m(\xi, t + s) = \Phi_m(\Phi_m(\xi, s), t)$. For all $x = (m, \xi) \in E$, define the deterministic exit time from E by

$$t^*(x) = \inf\{t > 0 \text{ such that } \Phi_m(\xi, t) \in \partial E_m\}.$$

Here and throughout, we use the convention that $\inf \emptyset = +\infty$.

- For all $m \in M$, the jump rate $\lambda_m: \bar{E}_m \rightarrow \mathbb{R}^+$ is measurable and satisfies the following condition:

$$\text{for all } (m, \xi) \in E, \text{ there exists } \varepsilon > 0 \text{ such that } \int_0^\varepsilon \lambda_m(\Phi_m(\xi, t)) dt < +\infty.$$

- For all $m \in M$, Q_m is a Markov kernel on $(\mathcal{B}(\bar{E}), \bar{E}_m)$ which satisfies the following condition:

$$\text{for all } \xi \in \bar{E}_m, \quad Q_m(E \setminus \{(m, \xi)\}, \xi) = 1.$$

From these characteristics, it can be shown (see [5, p. 57]) that there exists a filtered probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, (\mathbb{P}_x)_{x \in E})$ on which a process $(X_t)_{t \in \mathbb{R}^+}$ is defined. Its motion, starting from a point $x \in E$, may be constructed as follows. Let T_1 be a nonnegative random variable with survival function

$$\mathbb{P}_x(T_1 > t) = \begin{cases} e^{-\Lambda(x, t)} & \text{if } 0 \leq t < t^*(x), \\ 0 & \text{if } t \geq t^*(x), \end{cases}$$

where, for $x = (m, \xi) \in E$ and $t \in [0, t^*(x)]$,

$$\Lambda(x, t) = \int_0^t \lambda_m(\Phi_m(\xi, s)) ds.$$

We then choose an E -valued random variable Z_1 with distribution $Q_m(\cdot, \Phi_m(\xi, T_1))$. The trajectory of X_t for $t \leq T_1$ is

$$X_t = \begin{cases} (m, \Phi_m(\xi, t)) & \text{if } t < T_1, \\ Z_1 & \text{if } t = T_1. \end{cases}$$

Starting from the point $X_{T_1} = Z_1$, we select the next inter-jump time $T_2 - T_1$ and the next post-jump location Z_2 in a similar way.

Davis showed (see [5]) that the process so defined is a strong Markov process $(X_t)_{t \geq 0}$ with jump times $(T_n)_{n \in \mathbb{N}}$ (with $T_0 = 0$). The process $(\Theta_n)_{n \in \mathbb{N}} = (Z_n, T_n)_{n \in \mathbb{N}}$, where $Z_n = X_{T_n}$ is the post-jump location and T_n is the n th jump time, is clearly a discrete-time Markov chain. Besides, we denote by $S_n = T_n - T_{n-1}$ and $S_0 = 0$ the inter-jump times.

The following assumption about the jump times is standard (see, for example, [5, Section 24]).

Assumption 2.1. For all $(x, t) \in E \times \mathbb{R}^+$, $\mathbb{E}_x[\sum_k \mathbf{1}_{\{T_k < t\}}] < +\infty$.

Assumption 2.1 implies that $T_k \rightarrow +\infty$ almost surely (a.s.) when $k \rightarrow +\infty$.

For notational convenience, any function h defined on E will be identified with its component functions h_m defined on E_m . Thus, we write

$$h(x) = h_m(\xi) \quad \text{when } x = (m, \xi) \in E.$$

We also define a generalized flow $\Phi: E \times \mathbb{R}^+ \rightarrow E$ such that

$$\Phi(x, t) = (m, \Phi_m(\xi, t)) \quad \text{when } x = (m, \xi) \in E.$$

2.2. Notation

For any function w in $B(\bar{E})$, we introduce

$$Qw(x) = \int_E w(y) Q(dy, x), \quad C_w = \sup_{x \in \bar{E}} |w(x)|,$$

and, for any Lipschitz continuous function w in $B(\bar{E})$, we denote its Lipschitz constant by

$$[w] = \sup_{x \neq y \in \bar{E}} \frac{|w(x) - w(y)|}{|x - y|},$$

with the convention that $1/\infty = 0$.

Remark 2.1. For $w \in B(\bar{E})$ and from the definition of the distance on E , we have $[w] = \sup_{m \in M} [w_m]$.

3. Exit time

For all $m \in M$, let U_m be a Borel subset of E_m and let $U = \{(m, \xi), m \in M, \xi \in U_m\}$. We are interested in the exit time from U , denoted by τ and given by

$$\tau = \inf\{s \geq 0 \text{ such that } X_s \notin U\}.$$

Denote by μ the distribution of the initial state of the process Z_0 . Since the present paper concerns numerical computations, the following assumption appears natural.

Assumption 3.1. *The process starts in U and eventually leaves it a.s., i.e. the support of μ is included in U and $\mathbb{P}_\mu(\tau < +\infty) = 1$.*

The aim of this paper is to provide approximation schemes for the survival function and moments of the process. Our method has a high practical interest because it will provide numerical approximations as soon as the process can be simulated. Our approach is based on a recursive computation using the underlying discrete-time Markov chain $(Z_n, T_n)_{n \in \mathbb{N}}$. Therefore, we will study $\tau \wedge T_N$ rather than τ for some $N \in \mathbb{N}$ called the computation horizon. Indeed, thanks to Assumption 2.1, when N goes to ∞ , we have

$$\tau \wedge T_N \rightarrow \tau \quad \mathbb{P}_\mu\text{-a.s.}$$

One may approximate τ by $\tau \wedge T_N$ if N is chosen such that $\mathbb{P}_\mu(\tau > T_N)$ is small enough (the choice of N will be discussed in Section 3.3) because the evolution of the process beyond T_N will have little impact on the law or the moments of the exit time. In the rest of this section we present the two problems we are interested in and describe the recursive methods we use to solve them.

Definition 3.1. Let us define $u^*(x)$ for all $x \in U$ to be the time for the flow starting from the point x to exit from U , i.e.

$$u^*(x) = \inf\{s \geq 0 \text{ such that } \Phi(x, s) \notin U\}.$$

We now introduce some technical assumptions that will be in force throughout the paper. The first three assumptions will be crucial, while the two last assumptions can be made without loss of generality.

Assumption 3.2. *The function u^* is*

- (a) *Lipschitz continuous,*
- (b) *bounded by C_{u^*} .*

Assumption 3.3. *For all $m \in M$, the set U_m is convex.*

Assumption 3.4. *For $\alpha > 0$, let $U^\alpha = \{x \in E \text{ such that } d(x, \partial U) \leq \alpha\}$. There exist $C > 0$ and $\beta > 0$ such that, for all $k \in \{0, \dots, N\}$, $\mathbb{P}_\mu(Z_k \in U^\alpha) \leq C\alpha^\beta$.*

Remark 3.1. Assumption 3.4 can be checked in most of the applications. We will see, in the examples developed in Section 5, how it can be derived quite generally when Z_k has a bounded density. Moreover, it could be replaced by the following assumption, similar to an hypothesis

introduced in [5, Section 24] and presented as quite general in applications: there exists $\varepsilon > 0$ such that, for all $x \in U$, $Q(U^\varepsilon, x) = 0$, where $U^\varepsilon = \{x \in E \text{ such that } d(x, \partial U) \leq \varepsilon\}$, i.e. for all $k \in \{0, \dots, N\}$, $P_\mu(Z_k \in U^\varepsilon) = 0$.

Assumption 3.5. *The process cannot go back to U once it has left it, i.e. for all $z \in U^c$, $P_z(\text{there exists } t \geq 0, X_t \in U) = 0$.*

Assumption 3.6. *The function t^* is bounded by C_{t^*} .*

In our discussion, Assumption 3.5 does not imply any loss of generality and Assumption 3.6 stems from Assumption 3.2(b). Indeed, if any of the two previous assumptions is not satisfied by the process $(X_t)_{t \in \mathbb{R}^+}$, we introduce the process *killed at time τ* , denoted by $(\tilde{X}_t)_{t \in \mathbb{R}^+}$ and defined by

$$\tilde{X}_t = \begin{cases} X_t & \text{for } t < \tau, \\ \Delta & \text{for } t \geq \tau, \end{cases}$$

where Δ denotes a cemetery state. The state space of the killed process is $\tilde{E} = U \cup \{\Delta\}$ and Assumption 3.5 is fulfilled since the killed process remains in Δ after leaving U . In addition, \tilde{t}^* , the deterministic exit time from \tilde{E} for the killed process, equals u^* , which is bounded and Lipschitz continuous according to Assumption 3.2.

3.1. Distribution

The first goal of this paper is to compute an approximation for the law of the exit time τ . More precisely, we intend to approximate $P_\mu(\tau > s \mid \tau \leq T_N)$ for $s > 0$.

Our approach is of huge practical interest because we will see that, after some initial computations, any value of the survival function of τ may be quickly obtained. More importantly, our approach is even flexible with respect to U in the sense that the survival function of the exit time τ' from a new set $U' \subset U$ will also be directly available (provided that Assumptions 3.2–3.5 are still fulfilled by U').

Definition 3.2. For all $s > 0$, define the sequences $(p_k(s))_{k \geq 0}$, $(q_k)_{k \geq 0}$, and $(r_k(s))_{k \geq 0}$ as follows:

$$\begin{aligned} p_k(s) &= P_\mu(\tau > s \mid \tau \leq T_k), \\ q_k &= P_\mu(\tau \leq T_k), \\ r_k(s) &= P_\mu(\{\tau > s\} \cap \{T_k < \tau \leq T_{k+1}\}). \end{aligned}$$

Remark 3.2. The conditional probability $p_k(s)$ does not exist when $q_k = 0$. We then choose to extend the sequence by setting $p_k(s) = 0$.

Our objective is to approximate $p_N(s)$, where N represents the computation horizon. The following proposition provides a recursion for the sequence $(p_k)_{k \leq N}$; note that p_N may be computed as soon as the sequences $(q_k)_{k \leq N}$ and $(r_k)_{k \leq N-1}$ are known.

Proposition 3.1. *Under Assumption 3.1, for all $k \in \mathbb{N}$ and $s > 0$, $p_0(s) = 0$ and*

$$p_{k+1}(s) = \begin{cases} \frac{p_k(s)q_k + r_k(s)}{q_{k+1}} & \text{if } q_{k+1} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. First, recall that $T_0 = 0$ so that we have $p_0 = 0$ since the process starts in U according to Assumption 3.1. Then, let $k \in \mathbb{N}$ such that $q_{k+1} \neq 0$ and note that $\{\tau \leq T_{k+1}\} = \{\tau \leq T_k\} \cup \{T_k < \tau \leq T_{k+1}\}$. Then we have

$$\begin{aligned} p_{k+1}(s) &= \frac{\mathbb{P}_\mu(\{\tau > s\} \cap \{\tau \leq T_{k+1}\})}{\mathbb{P}_\mu(\tau \leq T_{k+1})} \\ &= \frac{\mathbb{P}_\mu(\{\tau > s\} \cap \{\tau \leq T_k\}) + \mathbb{P}_\mu(\{\tau > s\} \cap \{T_k < \tau \leq T_{k+1}\})}{q_{k+1}} \\ &= \frac{p_k(s)q_k + r_k(s)}{q_{k+1}}, \end{aligned}$$

completing the proof.

Now, before turning to computations, let us present the second problem we are interested in.

3.2. Moments

Our second goal is to approximate the moments of the exit time from U , i.e. for all $j \in \mathbb{N}$, we are interested in $\mathbb{E}_\mu[\tau^j \mid \tau \leq T_N]$. This is a very classical problem and some results are already available. First, it is possible to use a Monte Carlo method, and we will point out why the method we propose is more efficient and flexible. Furthermore, Helmes *et al.* [8] introduced a numerical method for computing the moments of the exit time based on linear programming. Lasserre and Prieto-Rumeau [9] improved this method by using semidefinite positivity moment conditions. These methods are quite efficient, but they require an assumption related to the generator of the process which is generally not fulfilled by the PDMP. The method we introduce now is based on the use of the Markov chain $(\Theta_n)_{n \in \mathbb{N}} = (Z_n, T_n)_{n \in \mathbb{N}}$ associated to the continuous-time process $(X_t)_{t \in \mathbb{R}^+}$.

Definition 3.3. For all $j \in \mathbb{N}$, introduce the sequences $(p_{k,j})_{k \geq 0}$ and $(r_{k,j})_{k \geq 0}$ defined as follows:

$$p_{k,j} = \mathbb{E}_\mu[\tau^j \mid \tau \leq T_k], \quad r_{k,j} = \mathbb{E}_\mu[\tau^j \mathbf{1}_{\{T_k < \tau \leq T_{k+1}\}}].$$

Our objective is to approximate $p_{N,j}$, where N still represents the computation horizon. Similarly to the previous section, the sequence $(p_{k,j})_{k \leq N}$ satisfies a recursion whose parameters are the sequences $(q_k)_{k \leq N}$, previously introduced, and $(r_{k,j})_{k \leq N-1}$.

Proposition 3.2. Under Assumption 3.1, we have, for all $k, j \in \mathbb{N}$, $p_{0,j} = 0$ and

$$p_{k+1,j} = \begin{cases} \frac{p_{k,j}q_k + r_{k,j}}{q_{k+1}} & \text{if } q_{k+1} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Proof. The proof is similar to that of Proposition 3.1.

Before turning to the approximation method itself, let us discuss the crucial question of the computation horizon.

3.3. The computation horizon

In this subsection we study more precisely the construction of the process (X_t) in order to obtain some results concerning the jump times $(T_k)_{k \in \mathbb{N}}$. For this purpose, we introduce, in this section only, two additional hypotheses.

Assumption 3.7. *The jump rate λ is bounded by C_λ .*

Assumption 3.8. *There exists $\varepsilon > 0$ such that, for all $x \in E$, $Q(x, A_\varepsilon) = 1$, where $A_\varepsilon = \{x \in E \text{ such that } t^*(x) \geq \varepsilon\}$. Roughly speaking, the jumps cannot send the process too close to the boundary of E .*

Assumption 3.7 is satisfied in a large majority of applications; Assumption 3.8 is quite general too and was introduced in [5, Section 24].

Let $(\Omega, \mathcal{A}, \mathbb{P})$ be a probability space on which is defined a sequence $(\Pi_k)_{k \in \mathbb{N}}$ of independent random variables with uniform distribution on $[0; 1]$. Let $x = (m, \xi) \in E$ and $\omega \in \Omega$, and let us focus on the construction of the trajectory $\{X_t(\omega), t > 0\}$ of the process starting from the point x . Let

$$F(t, x) = \begin{cases} 1 & \text{if } t \leq 0, \\ \exp\left(-\int_0^t \lambda(m, \Phi_m(\xi, s)) ds\right) & \text{if } 0 \leq t < t^*(x), \\ 0 & \text{if } t \geq t^*(x). \end{cases}$$

It is the survival function of the first jump time T_1 . Define its generalized inverse by

$$\Psi(u, x) = \begin{cases} \inf\{t \geq 0: F(t, x) \leq u\}, \\ +\infty, \end{cases} \quad \text{if the above set is empty.}$$

Let $S_1(\omega) = T_1(\omega) = \Psi(\Pi_1(\omega), x)$ and, for all $t < T_1(\omega)$,

$$X_t(\omega) = (m, \Phi_m(\xi, t)).$$

If $T_1(\omega) < +\infty$, choose X_{T_1} with distribution $Q(\cdot, \Phi_m(\xi, T_1))$. Assume that the trajectory is constructed until time T_k . If $T_k(\omega) < +\infty$, let

$$S_{k+1}(\omega) = \Psi(\Pi_k(\omega), X_{T_k}), \quad T_{k+1}(\omega) = T_k(\omega) + S_{k+1}(\omega).$$

If $T_{k+1}(\omega) < +\infty$, choose $X_{T_{k+1}}$ with distribution $Q(\cdot, \Phi_{m_{T_k}}(\xi_{T_k}, S_{k+1}))$. The trajectory is finally constructed by induction.

With the same notation as above, we state the following lemma.

Lemma 3.1. *Let H be a survival function such that, for all $t \in \mathbb{R}$ and all $x \in E$, $H(t) \leq F(t, x)$. There exists a sequence of independent random variables $(\tilde{S}_k)_{k \in \mathbb{N}}$ with distribution H and such that, for all $K \in \mathbb{R}$ and $N \in \mathbb{N}$,*

$$\mathbb{P}_\mu(T_N < K) \leq \mathbb{P}_\mu(\tilde{T}_N < K),$$

where $\tilde{T}_N = \sum_{k=0}^N \tilde{S}_k$.

Proof. Let H be such a survival function, and let $\tilde{\Psi}$ be its generalized inverse, i.e.

$$\tilde{\Psi}(u) = \begin{cases} \inf\{t \geq 0: H(t) \leq u\}, \\ +\infty, \end{cases} \quad \text{if the above set is empty.}$$

The assumption made on H yields, for all $x \in E$, $\tilde{\Psi}(u) \leq \Psi(u, x)$. Let, for all $k \in \mathbb{N}$ and all $\omega \in \Omega$,

$$\tilde{S}_k(\omega) = \tilde{\Psi}(\Pi_k(\omega)).$$

Note that we are using the same Π_k as in the definition of S_k , allowing us to write $\tilde{S}_k \leq S_k$ a.s. and, therefore, $\tilde{T}_k \leq T_k$ a.s. The result follows.

Similarly to Davis [5, Section 33], we approximated τ by $\tau \wedge T_N$ since $\tau \wedge T_N \rightarrow \tau$ as $N \rightarrow +\infty$ thanks to Assumption 2.1. It is therefore necessary to choose N large enough such that $\mathbb{P}_\mu(T_N < \tau)$ is small. It is difficult to estimate this probability for a general process because the links between τ and the jump times are largely problem dependent. For instance, the geometry of U can be very complex. Therefore, N will generally be estimated through simulations. Indeed, we can compute $\mathbb{P}_\mu(T_N < \tau)$ for some fixed N thanks to a Monte Carlo method and increase the value of N until this probability becomes small enough. However, we introduce another method to bound this probability that may prove useful in applications. First, note that, for any $K > 0$,

$$\{T_N < \tau\} \subset \{T_N < K\} \cup \{\tau > K\}.$$

This implies that

$$\mathbb{P}_\mu(T_N < \tau) \leq \mathbb{P}_\mu(T_N < K) + \mathbb{P}_\mu(\tau > K).$$

This will prove especially useful whenever τ is bounded, which happens quite often in applications, because there exists a K such that $\mathbb{P}_\mu(\tau > K) = 0$. When τ is not bounded, it is sometimes possible to obtain K such that $\mathbb{P}_\mu(\tau > K)$ is small.

Example 3.1. (*A crack propagation model.*) We adapt here an example studied by Chiquet and Limnios [4], which models a crack propagation. Here Y_t is a real-valued process representing the crack size and satisfying

$$Y_0 > 0, \quad \dot{Y}_t = A_t Y_t \quad \text{for all } t \geq 0,$$

where A_t is a Markov process with state space $\{\alpha, \beta\}$, $0 < \alpha \leq \beta$. We are interested in the time τ before the crack size reaches a critical size y_c . Consider the PDMP $X_t = (A_t, Y_t)$, where A_t represents the mode at time t . It is possible to bound the exit time by considering the slowest flow: we clearly have, for all $t \geq 0$, $Y_t \geq Y_0 e^{\alpha t}$ and, thus,

$$\mathbb{P}_\mu\left(\tau > \frac{1}{\alpha} \ln\left(\frac{y_c}{Y_0}\right)\right) = 0.$$

We now intend to bound $\mathbb{P}_\mu(T_N < K)$ for a fixed $K > 0$. Let

$$H(t) = \begin{cases} 1 & \text{if } t \leq 0, \\ e^{-C_\lambda t} & \text{if } 0 \leq t < \varepsilon, \\ 0 & \text{if } t \geq \varepsilon. \end{cases}$$

Distribution H represents, roughly speaking, the worst distribution of the inter-jump times S_k in the sense that it is the distribution that gives the most frequent jumps. Indeed, denote by F_k the survival function of S_k . We have $H \leq F_k$ for all $k \in \mathbb{N}$. Therefore, Lemma 3.1 provides a random variable $\tilde{T}_N = \sum_{k=0}^N \tilde{S}_k$, where the \tilde{S}_k are independent and have survival function H , such that

$$\mathbb{P}_\mu(T_N < K) \leq \mathbb{P}_\mu(\tilde{T}_N < K).$$

We now bound $\mathbb{P}_\mu(\tilde{T}_N < K)$. Standard computations yield $\mathbb{E}_\mu[\tilde{T}_N] = Nm$ and $\text{var}_\mu[\tilde{T}_N] = N\sigma^2$, where

$$m := \mathbb{E}_\mu[\tilde{S}_1] = \frac{1}{C_\lambda}(1 - e^{-C_\lambda \varepsilon}),$$

$$\sigma^2 := \text{var}_\mu[\tilde{S}_1] = \frac{1}{C_\lambda^2}(1 - 2C_\lambda \varepsilon e^{-C_\lambda \varepsilon} - e^{-2C_\lambda \varepsilon}).$$

Assume now that N is such that $Nm > K$, and note that

$$\mathbb{P}_\mu(\tilde{T}_N < K) \leq \mathbb{P}_\mu(|\tilde{T}_N - \mathbb{E}_\mu[\tilde{T}_N]| > \mathbb{E}_\mu[\tilde{T}_N] - K).$$

Tchebychev's inequality yields

$$\mathbb{P}_\mu(\tilde{T}_N < K) \leq \frac{N\sigma^2}{(Nm - K)^2}.$$

The right-hand side term goes to 0 when N goes to ∞ .

Finally, when τ is bounded with a high probability and when Assumptions 3.7 and 3.8 are fulfilled, we are able to choose N a priori such that $\mathbb{P}_\mu(T_N < \tau)$ is small. These conditions are satisfied in a large class of applications.

4. Approximation scheme

4.1. The quantization algorithm

First, we describe the quantization procedure for a random variable and recall some important properties that will be used in the sequel. There exists extensive literature on quantization methods for random variables and processes. We do not pretend to present here an exhaustive panorama of these methods. However, the interested reader is referred to, e.g. [1], [7], [10], and the references therein. Consider X , an \mathbb{R}^q -valued random variable such that $\|X\|_p < \infty$, where $\|X\|_p$ denotes the L_p -norm of X , i.e. $\|X\|_p = (\mathbb{E}[|X|^p])^{1/p}$.

Let K be a fixed integer. The optimal L_p -quantization of the random variable X consists in finding the best possible L_p -approximation of X by a random vector \hat{X} taking at most K values: $\hat{X} \in \{x^1, \dots, x^K\}$. This procedure consists of the following two steps.

1. Find a finite weighted grid $\Gamma \subset \mathbb{R}^q$ with $\Gamma = \{x^1, \dots, x^K\}$.
2. Set $\hat{X} = \hat{X}^\Gamma$, where $\hat{X}^\Gamma = \text{proj}_\Gamma(X)$, a Borel nearest-neighbour projection on Γ .

The asymptotic properties of the L_p -quantization are given by the following result; see, e.g. [10].

Theorem 4.1. *If $\mathbb{E}[|X|^{p+\eta}] < +\infty$ for some $\eta > 0$ then we have*

$$\lim_{K \rightarrow \infty} K^{p/q} \min_{|\Gamma| \leq K} \|X - \hat{X}^\Gamma\|_p^p = J_{p,q} \int |h|^{q/(q+p)}(u) \, du,$$

where the law of X is $P_X(du) = h(u)\lambda_q(du) + \nu$ with $\nu \perp \lambda_q$, $J_{p,q}$ a constant, and λ_q the Lebesgue measure in \mathbb{R}^q .

Note that X needs to have finite moments up to the order $p + \eta$ to ensure the above convergence. There exists a similar procedure for the optimal quantization of a Markov chain $\{X_k\}_{k \in \mathbb{N}}$. There are two approaches to provide the quantized approximation of a Markov chain. The first approach, based on the quantization at each time k of the random variable X_k , is called the *marginal quantization*. The second approach, which enhances the preservation of the Markov property, is called the *Markovian quantization*. Note that, for the latter, the quantized Markov process is not homogeneous. These two methods are described in detail in [10, Section 3]. In this work, we use the marginal quantization approach for simplicity reasons.

Our approximation methods are based on the quantization of the underlying discrete-time Markov chain $(\Theta_k)_{k \leq N} = (Z_k, T_k)_{k \leq N}$. The quantization algorithm provides, for each time

step $0 \leq k \leq N$, a finite grid Γ_k of $E \times \mathbb{R}^+$ as well as the transition matrices $(\widehat{Q}_k)_{0 \leq k \leq N-1}$ from Γ_k to Γ_{k+1} . Let $p \geq 1$ such that, for all $k \leq N$, Z_k and T_k have finite moments at least up to order p and let proj_{Γ_k} be the nearest-neighbour projection from $E \times \mathbb{R}^+$ onto Γ_k . The quantized process $(\widehat{\Theta}_k)_{k \leq N} = (\widehat{Z}_k, \widehat{T}_k)_{k \leq N}$ with value for each k in the finite grid Γ_k of $E \times \mathbb{R}^+$ is then defined by

$$(\widehat{Z}_k, \widehat{T}_k) = \text{proj}_{\Gamma_k}(Z_k, T_k).$$

In practice, we begin with the computation of the quantization grids, which merely requires us to be able to simulate the process. These grids are computed only once and may be stored offline. Our schemes are then based on the following simple idea: we replace the process by its quantized approximation within the different recursions. The results are obtained in a very simple way since the quantized process has finite state space.

Remark 4.1. In addition, we recall a technical property of the quantization algorithm proved by Bouton and Pagès in [3]: the quantized process evolves within the convex hull of the support of the law of the original process. Therefore, and it will be required below, it follows from Assumption 3.3 that if $Z_k \in U$ a.s. for some $k \in \{0, \dots, N\}$ then $\widehat{Z}_k \in U$ a.s.

4.2. Approximation scheme of the distribution and proof of convergence

We already noted in Proposition 3.1 that $p_N(s) = \mathbb{P}_\mu(\tau > s \mid \tau \leq T_N)$ may be computed as soon as the sequences $(q_k)_{k \leq N}$ and $(r_k)_{k \leq N-1}$ are known. Therefore, we find expressions of these sequences depending on the Markov chain $(Z_k, T_k)_{k \leq N}$, which we replace by the quantized process $(\widehat{Z}_k, \widehat{T}_k)_{k \leq N}$ in order to define their quantized approximations $(\widehat{q}_k)_{k \leq N}$ and $(\widehat{r}_k)_{k \leq N-1}$.

First, note that $\{T_k < \tau\} = \{Z_k \in U\}$ and $\{\tau \leq T_k\} = \{Z_k \notin U\}$ thanks to Assumption 3.5. Moreover, on $\{Z_k \in U, Z_{k+1} \notin U\}$, we have $\tau = (T_k + u^*(Z_k)) \wedge T_{k+1}$ a.s., where $u^*(x)$ is the deterministic exit time from U starting from the point x (see Definition 3.1), and we have

$$q_k = \mathbb{E}_\mu[\mathbf{1}_{U^c}(Z_k)], \quad r_k(s) = \mathbb{E}_\mu[\mathbf{1}_{\{(T_k + u^*(Z_k)) \wedge T_{k+1} > s\}} \mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1})]. \quad (4.1)$$

The above equations are crucial in our discussion and, from now on, we will use them without referring to Assumption 3.5.

Before turning to the approximation scheme itself, let us state some properties of the sequence $(q_k)_{k \leq N}$ that will be important in the following proofs. Indeed, the sequence $(q_k)_k$ increases since $\{\tau \leq T_k\} \subset \{\tau \leq T_{k+1}\}$ for all $k \leq N-1$. Moreover, note that $q_0 = 0$ and $\lim_{n \rightarrow +\infty} q_n = 1$ thanks to Assumption 3.1. Therefore, there exists an index, denoted by $\tilde{k} \geq 1$, such that

- for all $k < \tilde{k}$, we have $q_k = 0$,
- for all $k \geq \tilde{k}$, we have $q_k > 0$.

We denote by $\tilde{q} = q_{\tilde{k}}$ the first positive value of the sequence so that $q_k \geq \tilde{q}$ for all $k \geq \tilde{k}$. Then we obtain the following definition.

Definition 4.1. Let

$$\tilde{k} = \inf\{k \geq 0 \text{ such that } q_k > 0\}, \quad \tilde{q} = q_{\tilde{k}},$$

i.e. \tilde{q} is the first strictly positive value of the sequence $(q_k)_{k \in \{0, \dots, N\}}$.

We now naturally define the quantized approximations of the previous sequences.

Definition 4.2. For all $s > 0$, define the sequences $(\widehat{q}_k)_{k \in \{0, \dots, N\}}$ and $(\widehat{r}_k)_{k \in \{0, \dots, N-1\}}$ by

$$\begin{aligned}\widehat{q}_k &= \mathbf{E}_\mu[\mathbf{1}_{U^c}(\widehat{Z}_k)], \\ \widehat{r}_k(s) &= \mathbf{E}_\mu[\mathbf{1}_{\{(\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1} > s\}} \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1})].\end{aligned}$$

It is important to note that both \widehat{q}_k and $\widehat{r}_k(s)$ may be computed easily from the quantization algorithm. Indeed, we have

$$\begin{aligned}\widehat{q}_k &= \sum_{\substack{\theta=(z,t) \in \Gamma_k \\ z \notin U}} \mathbf{P}(\widehat{\Theta}_k = \theta), \\ \widehat{r}_k(s) &= \sum_{\substack{\theta=(z,t) \in \Gamma_k \\ z \in U}} \sum_{\substack{\theta'=(z',t') \in \Gamma_{k+1} \\ z' \notin U}} \mathbf{1}_{\{(t+u^*(z)) \wedge t' > s\}} \mathbf{P}(\widehat{\Theta}_k = \theta) \widehat{Q}_k(\theta; \theta').\end{aligned}$$

Recall from Proposition 3.1 that the sequence $(p_k)_{k \leq N}$ satisfies a recursion that depends on the two parameters $(q_k)_{k \leq N}$ and $(r_k)_{k \leq N-1}$, which we are now able to approximate. Hence, replacing them by their quantized approximations within the same recursion leads to a new sequence, denoted by $(\widehat{p}_k)_{k \leq N}$. The rest of this section is dedicated to the proof of the convergence of $(\widehat{p}_k)_{k \leq N}$ towards $(p_k)_{k \leq N}$. This convergence is far from trivial because, on the one hand, the definitions of the sequences $(q_k)_{k \leq N}$ and $(r_k)_{k \leq N-1}$ contain many indicator functions that are not Lipschitz continuous and, on the other hand, the recursive function giving p_{k+1} from p_k , q_k , q_{k+1} , and r_k is not Lipschitz continuous either.

Definition 4.3. For all $s > 0$ and all $k \in \{0, \dots, N-1\}$, let $\widehat{p}_0(s) = 0$ and

$$\widehat{p}_{k+1}(s) = \begin{cases} \frac{\widehat{p}_k(s)\widehat{q}_k + \widehat{r}_k(s)}{\widehat{q}_{k+1}} & \text{if } \widehat{q}_{k+1} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

The two following propositions will be necessary to prove the convergence of the approximation scheme. They respectively state the convergences of $(\widehat{q}_k)_{k \leq N}$ and $(\widehat{r}_k)_{k \leq N-1}$ towards $(q_k)_{k \leq N}$ and $(r_k)_{k \leq N-1}$.

Proposition 4.1. Under Assumptions 3.4 and 3.5, for all $k \in \{0, \dots, N\}$, \widehat{q}_k converges towards q_k when the quantization error $\|\Theta_k - \widehat{\Theta}_k\|_p$ goes to 0. More precisely, the error is bounded by

$$|q_k - \widehat{q}_k| \leq C^{p/(p+\beta)} \left(\left(\frac{\beta}{p} \right)^{p/(p+\beta)} + \left(\frac{p}{\beta} \right)^{\beta/(p+\beta)} \right) \|Z_k - \widehat{Z}_k\|_p^{p\beta/(p+\beta)},$$

where C and β are defined in Assumption 3.4.

Proof. For all $k \in \{0, \dots, N\}$, (4.1) yields

$$|q_k - \widehat{q}_k| = |\mathbf{E}_\mu[\mathbf{1}_U(Z_k) - \mathbf{1}_U(\widehat{Z}_k)]|.$$

The difference between the indicator functions is nonzero if and only if Z_k and \widehat{Z}_k are on either side of ∂U . Therefore, in this case, for all $\alpha > 0$, if $|Z_k - \widehat{Z}_k| \leq \alpha$ then $d(Z_k, \partial U) \leq \alpha$. Hence, either $|Z_k - \widehat{Z}_k| > \alpha$ or $Z_k \in U^\alpha$. The Markov inequality and Assumption 3.4 yield

$$\begin{aligned}\mathbf{E}_\mu |\mathbf{1}_U(Z_k) - \mathbf{1}_U(\widehat{Z}_k)| &\leq \mathbf{P}_\mu(|Z_k - \widehat{Z}_k| > \alpha) + \mathbf{P}_\mu(Z_k \in U^\alpha) \\ &\leq \frac{\|Z_k - \widehat{Z}_k\|_p^p}{\alpha^p} + C\alpha^\beta.\end{aligned}$$

This bound reaches a minimum when

$$\alpha = \left(\frac{p \|Z_k - \widehat{Z}_k\|_p^p}{\beta C} \right)^{1/(p+\beta)},$$

and the result follows.

Proposition 4.2. *Under Assumptions 3.2(a), 3.4, and 3.5, for all $k \in \{0, \dots, N-1\}$ and almost every $s > 0$ with respect to the Lebesgue measure on \mathbb{R} ,*

$$\widehat{r}_k(s) \rightarrow r_k(s)$$

when the quantization errors $\|\Theta_l - \widehat{\Theta}_l\|_p$ for $l \in \{k, k+1\}$ go to 0.

Proof. Let $k \in \{0, \dots, N-1\}$ and $s > 0$. Equation (4.1) yields

$$|r_k(s) - \widehat{r}_k(s)| \leq A + B,$$

where

$$\begin{aligned} A &= |\mathbf{E}_\mu[\mathbf{1}_{\{(T_k + u^*(Z_k)) \wedge T_{k+1} > s\}} - \mathbf{1}_{\{(\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1} > s\}}] \mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1})]|, \\ B &= |\mathbf{E}_\mu[\mathbf{1}_{\{(\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1} > s\}} (\mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1}) - \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1}))]|. \end{aligned}$$

In the A term, we crudely bound $\mathbf{1}_U(Z_k)$ and $\mathbf{1}_{U^c}(Z_{k+1})$ by 1 and turn to the difference between the two indicator functions. This difference is nonzero if and only if $(T_k + u^*(Z_k)) \wedge T_{k+1}$ and $(\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1}$ are on either side of s , implying that they both belong to $[s - \eta; s + \eta]$, where $\eta = |(T_k + u^*(Z_k)) \wedge T_{k+1} - (\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1}|$. Then we have

$$|\mathbf{1}_{\{(T_k + u^*(Z_k)) \wedge T_{k+1} > s\}} - \mathbf{1}_{\{(\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1} > s\}}| \leq \mathbf{1}_{\{|(T_k + u^*(Z_k)) \wedge T_{k+1} - s| \leq \eta\}}$$

so that

$$A \leq \mathbf{P}_\mu(|(T_k + u^*(Z_k)) \wedge T_{k+1} - s| \leq \eta).$$

The following discussion consists in noting that either η and the probability that $(T_k + u^*(Z_k)) \wedge T_{k+1}$ belongs to the interval $[s - \eta; s + \eta]$ are small, or η is large, but this happens with a small probability too when the quantization error goes to 0. For all $\alpha > 0$, we have

$$\begin{aligned} A &\leq \mathbf{P}_\mu(|(T_k + u^*(Z_k)) \wedge T_{k+1} - s| \leq \eta, \eta \leq \alpha) + \mathbf{P}_\mu(\eta > \alpha) \\ &\leq \mathbf{P}_\mu(|(T_k + u^*(Z_k)) \wedge T_{k+1} - s| \leq \alpha) + \mathbf{P}_\mu(\eta > \alpha) \\ &\leq |\varphi_k(s + \alpha) - \varphi_k(s - \alpha)| + \frac{\|\eta\|_p^p}{\alpha^p}, \end{aligned}$$

where φ_k denotes the distribution function of $(T_k + u^*(Z_k)) \wedge T_{k+1}$. Let $\varepsilon > 0$, and assume that s is not an atom of this distribution, so that there exists $\alpha_1 > 0$ such that $|\varphi_k(s + \alpha_1) - \varphi_k(s - \alpha_1)| \leq \varepsilon$. Besides, thanks to Assumption 3.2(a), the Lipschitz continuity condition on u^* , we have $\eta \leq |T_k - \widehat{T}_k| + [u^*] \|Z_k - \widehat{Z}_k\| + |T_{k+1} - \widehat{T}_{k+1}|$. Moreover, since the quantization error goes to 0, we may assume that $\|\eta\|_p \leq \alpha_1 \varepsilon^{1/p}$. Setting $\alpha = \alpha_1$ in the previous computations yields

$$A \leq |\varphi_k(s + \alpha_1) - \varphi_k(s - \alpha_1)| + \frac{\|\eta\|_p^p}{\alpha_1^p} \leq 2\varepsilon.$$

Note that the set of atoms of the distribution function of $(T_k + u^*(Z_k)) \wedge T_{k+1}$ is at most countable, so the previous discussion is true for almost every $s > 0$ with respect to the Lebesgue measure. Let us now bound the B term:

$$\begin{aligned} B &\leq \mathbb{E}_\mu |\mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1}) - \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1})| \\ &\leq \mathbb{E}_\mu [\mathbf{1}_{U^c}(Z_{k+1}) |1_U(Z_k) - 1_U(\widehat{Z}_k)|] + \mathbb{E}_\mu [1_U(\widehat{Z}_k) |1_{U^c}(Z_{k+1}) - 1_{U^c}(\widehat{Z}_{k+1})|] \\ &\leq |q_k - \widehat{q}_k| + |q_{k+1} - \widehat{q}_{k+1}|, \end{aligned}$$

which goes to 0 thanks to Proposition 4.1.

The convergence of the approximation scheme of the distribution of the exit time is now a straightforward consequence of the following proposition.

Proposition 4.3. *We assume that Assumptions 3.1, 3.3, 3.4, and 3.5 hold. Let $(\sigma_k)_{k \leq N-1}$ and $(\widehat{\sigma}_k)_{k \leq N-1}$ be two sequences of $[0, 1]$ -valued real numbers. Let $(\pi_k)_{0 \leq k \leq N}$ and $(\widehat{\pi}_k)_{0 \leq k \leq N}$ be defined as follows: $\pi_0 = \widehat{\pi}_0 = 0$,*

$$\begin{aligned} \pi_{k+1} &= \begin{cases} \frac{\pi_k q_k + \sigma_k}{q_{k+1}} & \text{if } q_{k+1} \neq 0, \\ 0 & \text{otherwise,} \end{cases} \\ \widehat{\pi}_{k+1} &= \begin{cases} \frac{\widehat{\pi}_k \widehat{q}_k + \widehat{\sigma}_k}{\widehat{q}_{k+1}} & \text{if } \widehat{q}_{k+1} \neq 0, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

For $0 \leq k \leq N$, if the quantization error is such that, for all $l \leq k$,

$$C^{p/(p+\beta)} \left(\left(\frac{\beta}{p} \right)^{p/(p+\beta)} + \left(\frac{p}{\beta} \right)^{\beta/(p+\beta)} \right) \|Z_l - \widehat{Z}_l\|_p^{p\beta/(p+\beta)} \leq \frac{1}{2} \tilde{q},$$

then

$$\begin{aligned} |\pi_k - \widehat{\pi}_k| &\leq \frac{2}{\tilde{q}} (\pi^{\sup} |q_{k-1} - \widehat{q}_{k-1}| + |\pi_{k-1} - \widehat{\pi}_{k-1}| + |\sigma_{k-1} - \widehat{\sigma}_{k-1}|) \\ &\quad + \frac{2(\pi^{\sup} + 1)}{\tilde{q}^2} |q_k - \widehat{q}_k|, \end{aligned}$$

where $\pi^{\sup} = \max_{0 \leq k \leq N} \pi_k$.

Proof. The difficulty with proving this result lies in the fact that the recursive function giving π_{k+1} from π_k , q_k , q_{k+1} , and σ_k is not Lipschitz continuous because of the division by q_{k+1} . To overcome this drawback, we will use the strictly positive lower bound for q_k described earlier. Indeed, recall from Definition 4.1 that there exists a step \tilde{k} such that $q_k \geq \tilde{q} > 0$ for all $k \geq \tilde{k}$ and $q_k = 0$ for all $k < \tilde{k}$. What is more, a similar bound will be derived for the quantized values \widehat{q}_k thanks to the convergence of \widehat{q}_k towards q_k .

We now prove by induction that $\widehat{\pi}_k$ converges towards π_k . First, we have $\widehat{\pi}_0 = \pi_0 = 0$. Then, let $k \in \{1, \dots, N\}$.

If $k < \tilde{k}$ then $q_k = 0$ and Assumption 3.3 yields $\widehat{q}_k = 0$ too. Indeed, $q_k = 0$ means that $Z_k \in U$ a.s. Since U is a convex set, Remark 4.1 implies that $\widehat{Z}_k \in U$ a.s. too. In other words, $\widehat{q}_k = 0$. Finally, from the definitions, we have $\pi_k = \widehat{\pi}_k = 0$.

If $k \geq \tilde{k}$ then $q_k \geq \tilde{q} > 0$. In order to bound the error between π_k and $\widehat{\pi}_k$, it is indeed necessary to have a strictly positive lower bound for q_k because of the division by q_k within the

recursion. Now we need to obtain the same kind of bound for \widehat{q}_k . This can be achieved thanks to Proposition 4.1, giving the convergence of \widehat{q}_k towards q_k . Indeed, assume from now on that the number of points in the quantization grids is large enough such that the quantization error is sufficiently small to ensure that, for all $j = \widehat{k}, \dots, N$, $|q_j - \widehat{q}_j| \leq \frac{1}{2}\tilde{q}$. Hence, the required lower bound is $\widehat{q}_k \geq \frac{1}{2}\tilde{q} > 0$. Therefore,

$$\begin{aligned}
|\pi_k - \widehat{\pi}_k| &\leq \left| \frac{\pi_{k-1}q_{k-1} + \sigma_{k-1}}{q_k} - \frac{\widehat{\pi}_{k-1}\widehat{q}_{k-1} + \widehat{\sigma}_{k-1}}{\widehat{q}_k} \right| \\
&\leq \frac{\pi_{k-1}}{\widehat{q}_k} |q_{k-1} - \widehat{q}_{k-1}| + \frac{\widehat{q}_{k-1}}{\widehat{q}_k} |\pi_{k-1} - \widehat{\pi}_{k-1}| + \frac{1}{\widehat{q}_k} |\sigma_{k-1} - \widehat{\sigma}_{k-1}| \\
&\quad + |\pi_{k-1}q_{k-1} + \sigma_{k-1}| \frac{|q_k - \widehat{q}_k|}{q_k \widehat{q}_k} \\
&\leq \frac{\pi^{\sup}}{\widehat{q}_k} |q_{k-1} - \widehat{q}_{k-1}| + \frac{1}{\widehat{q}_k} |\pi_{k-1} - \widehat{\pi}_{k-1}| + \frac{1}{\widehat{q}_k} |\sigma_{k-1} - \widehat{\sigma}_{k-1}| \\
&\quad + (\pi^{\sup} + 1) \frac{|q_k - \widehat{q}_k|}{q_k \widehat{q}_k} \\
&\leq \frac{2}{\tilde{q}} (\pi^{\sup} |q_{k-1} - \widehat{q}_{k-1}| + |\pi_{k-1} - \widehat{\pi}_{k-1}| + |\sigma_{k-1} - \widehat{\sigma}_{k-1}|) \\
&\quad + \frac{2(\pi^{\sup} + 1)}{\tilde{q}^2} |q_k - \widehat{q}_k|,
\end{aligned}$$

where $\pi^{\sup} = \max_{0 \leq k \leq N} \pi_k$.

Remark 4.2. Note that a bound for the rate of convergence of $\widehat{\pi}_k$ towards π_k may be obtained as soon as a bound for the rate of convergence of $\widehat{\sigma}_k$ towards σ_k and an upper bound for the sequence $(\pi_k)_{0 \leq k \leq N}$ are available.

We now state one of our main results, namely the convergence of the approximation scheme of the distribution of the exit time.

Theorem 4.2. *Under Assumptions 3.1, 3.2(a), 3.3, 3.4, and 3.5, for all $k \in \{0, \dots, N\}$ and almost every $s > 0$ with respect to the Lebesgue measure on \mathbb{R} ,*

$$\widehat{p}_k(s) \rightarrow p_k(s)$$

when the quantization errors $\|\Theta_j - \widehat{\Theta}_j\|_p$ for $j \in \{0, \dots, k\}$ go to 0.

Proof. Let $s > 0$ such that $(\widehat{r}_k(s))_k$ converges towards $(r_k(s))_k$ and apply Proposition 4.3 with $(\sigma_k)_k = (r_k(s))_k$ and $(\widehat{\sigma}_k)_k = (\widehat{r}_k(s))_k$ so that $(\pi_k)_k = (p_k(s))_k$ and $(\widehat{\pi}_k)_k = (\widehat{p}_k(s))_k$. Finally, note that $(p_k(s))_k$ is bounded by 1.

Remark 4.3. It may be useful to note that, although it will be crucial in the moments approximation scheme, the boundedness condition on u^* (Assumption 3.2(b)) was unnecessary in this section. Hence, the distribution approximation can be achieved without this hypothesis.

We now obtain an easily computable approximation for the survival function of the exit time. Let us now consider its moments. Of course, they can be derived from the distribution, but we present in the following subsection a method to approximate them directly. An important advantage of this method will be to provide a bound for the rate of convergence.

4.3. Approximation scheme of the moments and rate of convergence

Similarly to the distribution, the moments can be approximated thanks to the quantization of the process $(\Theta_k)_{k \leq N} = (Z_k, T_k)_{k \leq N}$. However, it is important to stress that we will be able to derive a rate of convergence for our approximation scheme. We note from Proposition 3.2 that, similarly to the case of the distribution, $p_{N,j} = \mathbb{E}_\mu[\tau^j \mid \tau \leq T_N]$ can be computed as soon as the sequences $(q_k)_{k \leq N}$ and $(r_{k,j})_{k \leq N-1}$ are known. The first sequence has already been approximated in the previous section, but we still need to find an expression for the second sequence, dependent on the Markov chain $(Z_k, T_k)_k$ to define its quantized approximation $(\widehat{r}_{k,j})_{k \leq N-1}$. Thanks to Assumption 3.5, the same arguments give

$$r_{k,j} = \mathbb{E}_\mu[((T_k + u^*(Z_k)) \wedge T_{k+1})^j \mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1})].$$

Hence, we can now naturally define the quantized approximation of the sequences $(r_{k,j})_{k \leq N-1}$ and $(p_{k,j})_{k \leq N}$.

Definition 4.4. For all $j \in \mathbb{N}$, define the sequence $(\widehat{r}_{k,j})_{k \in \{0, \dots, N-1\}}$ by

$$\widehat{r}_{k,j} = \mathbb{E}_\mu[((\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1})^j \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1})]$$

and the sequence $(\widehat{p}_{k,j})_{k \in \{0, \dots, N\}}$ by $\widehat{p}_{0,j} = 0$ and

$$\widehat{p}_{k+1,j} = \begin{cases} \frac{\widehat{p}_{k,j} \widehat{q}_k + \widehat{r}_{k,j}}{\widehat{q}_{k+1}} & \text{if } \widehat{q}_{k+1} \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

As for \widehat{q}_k and $\widehat{r}_k(s)$ defined in the previous section, $\widehat{r}_{k,j}$ may be computed easily from the quantization algorithm. Indeed, we have

$$\widehat{r}_{k,j} = \sum_{\substack{\theta=(z,t) \in \Gamma_k \\ z \in U}} \sum_{\substack{\theta'=(z',t') \in \Gamma_{k+1} \\ z' \notin U}} ((t + u^*(z)) \wedge t')^j \mathbb{P}(\widehat{\Theta}_k = \theta) \widehat{Q}_k(\theta; \theta').$$

The following proposition proves the convergence of $\widehat{r}_{k,j}$ towards $r_{k,j}$.

Proposition 4.4. Under Assumptions 3.2(a), 3.4, 3.5, and 3.6, for all $k \in \{0, \dots, N-1\}$ and all $j \in \mathbb{N}$, $\widehat{r}_{k,j}$ converges towards $r_{k,j}$ when the quantization errors $\|\Theta_l - \widehat{\Theta}_l\|_p$ for $l \in \{k, k+1\}$ go to 0. More precisely, the error is bounded by

$$|r_{k,j} - \widehat{r}_{k,j}| \leq j((k+1)C_{t^*})^{j-1} (\|T_k - \widehat{T}_k\|_p + [u^*]\|Z_k - \widehat{Z}_k\|_p + \|T_{k+1} - \widehat{T}_{k+1}\|_p) \\ + ((k+1)C_{t^*})^j (|q_k - \widehat{q}_k| + |q_{k+1} - \widehat{q}_{k+1}|).$$

Proof. Let $k \in \{0, \dots, N-1\}$ and $j \in \mathbb{N}$. We have

$$|r_{k,j} - \widehat{r}_{k,j}| \leq A + B,$$

where

$$A = |\mathbb{E}_\mu[((T_k + u^*(Z_k)) \wedge T_{k+1})^j - ((\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1})^j \mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1})]|, \\ B = |\mathbb{E}_\mu[((\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1})^j (\mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1}) - \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1}))]|.$$

It follows from Assumption 3.6 that the inter-jump times S_i are a.s. bounded by C_{t^*} , so $T_i \leq iC_{t^*}$ a.s. and $(T_i + u^*(Z_i)) \wedge T_{i+1} \leq (i+1)C_{t^*}$ a.s. By Remark 4.1, these bounds are equally true for the quantized process $\widehat{T}_i \leq iC_{t^*}$ and $(\widehat{T}_i + u^*(\widehat{Z}_i)) \wedge \widehat{T}_{i+1} \leq \widehat{T}_{i+1} \leq (i+1)C_{t^*}$ a.s.

Let us first consider the term A . We crudely bound the indicator functions by 1. Moreover, define $\eta = |(T_k + u^*(Z_k)) \wedge T_{k+1} - (\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1}|$ and note that the function $x \rightarrow x^j$ is Lipschitz continuous on any set $[0, M]$ with Lipschitz constant jM^{j-1} . Then

$$A \leq \mathbb{E}_\mu[j((k+1)C_{t^*})^{j-1}\eta] \leq j((k+1)C_{t^*})^{j-1}\|\eta\|_p,$$

and thanks to Assumption 3.2(a), the Lipschitz continuity condition on u^* , we have

$$A \leq j((k+1)C_{t^*})^{j-1}(\|T_k - \widehat{T}_k\|_p + [u^*]\|Z_k - \widehat{Z}_k\|_p + \|T_{k+1} - \widehat{T}_{k+1}\|_p).$$

Moreover, the term B is bounded by

$$\begin{aligned} B &\leq ((k+1)C_{t^*})^j \mathbb{E}_\mu[|\mathbf{1}_U(Z_k) \mathbf{1}_{U^c}(Z_{k+1}) - \mathbf{1}_U(\widehat{Z}_k) \mathbf{1}_{U^c}(\widehat{Z}_{k+1})| \\ &\leq ((k+1)C_{t^*})^j (|q_k - \widehat{q}_k| + |q_{k+1} - \widehat{q}_{k+1}|). \end{aligned}$$

Using Proposition 4.1 completes the proof.

We may now state the other important results of our paper, namely the convergence of the approximation scheme of the moments of the exit time with a bound for the rate of convergence.

Theorem 4.3. *Under Assumptions 3.1, 3.2(a), 3.3, 3.4, 3.5, and 3.6, for all $k \in \{0, \dots, N\}$ and all $j \in \mathbb{N}$, $\widehat{p}_{k,j}$ converges towards $p_{k,j}$ when the quantization errors $\|\Theta_j - \widehat{\Theta}_j\|_p$ for $j \in \{0, \dots, k\}$ go to 0.*

More precisely, if the quantization error is such that, for all $l \leq k$,

$$C^{p/(p+q)} \left(\left(\frac{q}{p} \right)^{p/(p+q)} + \left(\frac{p}{q} \right)^{q/(p+q)} \right) \|Z_l - \widehat{Z}_l\|_p^{pq/(p+q)} \leq \frac{1}{2} \tilde{q},$$

then

$$\begin{aligned} |p_{k,j} - \widehat{p}_{k,j}| &\leq \frac{2}{\tilde{q}} ((NC_{t^*})^j |q_{k-1} - \widehat{q}_{k-1}| + |p_{k-1,j} - \widehat{p}_{k-1,j}| + |r_{k-1,j} - \widehat{r}_{k-1,j}|) \\ &\quad + \frac{2((NC_{t^*})^j + 1)}{\tilde{q}^2} |q_k - \widehat{q}_k|. \end{aligned}$$

Remark 4.4. The rate of convergence depends on the quantity \tilde{q} whose exact value might be unknown in some complex applications. In that case, it may still be approximated through Monte Carlo simulations (see the examples in Section 5). Nevertheless, Theorems 4.2 and 4.3 prove the convergence of our approximation schemes regardless of the value of \tilde{q} .

Proof of Theorem 4.3. Let $j \in \mathbb{N}$, and apply Proposition 4.3 with $(\sigma_k)_k = (r_{k,j})_k$ and $(\widehat{\sigma}_k)_k = (\widehat{r}_{k,j})_k$ such that $(\pi_k)_k = (p_{k,j})_k$ and $(\widehat{\pi}_k)_k = (\widehat{p}_{k,j})_k$. Finally, according to Remark 4.2, a bound for the rate of convergence is obtained since the sequence $(p_{k,j})_{0 \leq k \leq N}$ is bounded by

$$p_{k,j} = \mathbb{E}_\mu[\tau^j \mid \tau \leq T_k] \leq \mathbb{E}_\mu[T_k^j \mid \tau \leq T_k] \leq \mathbb{E}_\mu[(kC_{t^*})^j \mid \tau \leq T_k] \leq (kC_{t^*})^j \leq (NC_{t^*})^j.$$

This completes the proof.

5. Examples and numerical results

5.1. A Poisson process example

Let N_t be a Poisson process with parameter $\lambda = 1$, and let $Y_t = t + N_t$. Here $(Y_t)_{t \geq 0}$ is a PDMP with state space $E = \mathbb{R}$; the inter-jump times S_k have independent exponential distribution with parameter $\lambda = 1$; the flow is defined on $(\mathbb{R}^+)^2$ by $\Phi(x, t) = x + t$; and, finally, the post-jump locations satisfy, for all $x \in E$, $Q(\{x + 1\}, x) = 1$. An example of a trajectory of the process is represented in Figure 1. We are interested in the exit time problem for the process $(Y_t)_{t \geq 0}$. The study of this process is especially interesting because it is possible to compute the exact value of its distribution function in order to compare it with the numerical value given by our approximation scheme.

Let us turn now to the numerical simulations. Let $b = 10$, i.e. $U = (-\infty, 10)$. We may choose $N = 10$ since $Y_{T_N} = T_N + N_{T_N} = T_N + N \geq N$. Besides, it is clear that, for all $y \in (-\infty, 10)$, $u^*(y) = 10 - y$. Assumptions 3.2 and 3.3 are clearly satisfied and so is Assumption 3.4 thanks to the following lemma.

Lemma 5.1. *For all $\alpha > 0$ and all $k \in \{0, \dots, N\}$,*

$$\mathbb{P}_\mu(Z_k \in U^\alpha) \leq 2\alpha.$$

Proof. Since $Z_0 = 0$ a.s., $\mathbb{P}_\mu(Z_0 \in U^\alpha) = \mathbb{P}_\mu(Z_0 \in [10 - \alpha, 10 + \alpha]) = \mathbf{1}_{\{\alpha \geq 10\}} \leq \frac{1}{10}\alpha \leq 2\alpha$.

Now let $k \in \{1, \dots, N\}$. Denote by $f_{\gamma(k,1)}$ the density of the distribution $\gamma(k, 1)$, and let its bound be denoted by

$$C_k = \frac{1}{(k-1)!} \left(\frac{k-1}{e} \right)^{k-1}.$$

Since T_k has distribution $\gamma(k, 1)$, $Z_k = k + T_k$ has density $f_{Z_k}(\cdot) = f_{\gamma(k,1)}(\cdot - k)$, which is also bounded by C_k . Eventually, we have

$$\mathbb{P}_\mu(Z_k \in U^\alpha) = \mathbb{P}_\mu(Z_k \in [10 - \alpha, 10 + \alpha]) \leq 2C_k\alpha \leq 2\alpha.$$

Indeed, the sequence $(C_k)_k$ decreases so that, for all $k \in \{1, \dots, N\}$, $C_k \leq C_1 = 1$.

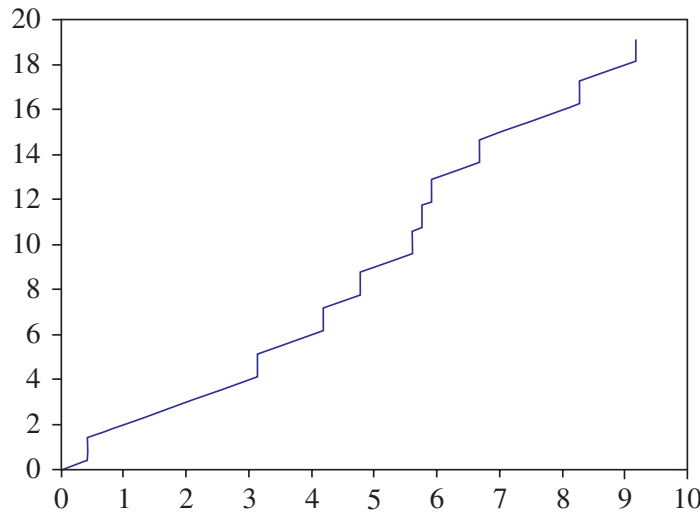


FIGURE 1: A trajectory of the process (Y_t) drawn until the 10th jump time.

TABLE 1: Simulation results for the mean exit time.

Number of points in the quantization grids	$\widehat{p}_{N,1}$	Relative error to 5.125 (%)
20	5.050	1.46
50	5.096	0.56
100	5.095	0.58
200	5.118	0.13
300	5.128	0.06
500	5.123	0.03

TABLE 2: Simulation results for the second moment.

Number of points in the quantization grids	$\widehat{p}_{N,2}$	Relative error to 27.5 (%)
20	26.66	3.05
50	27.20	1.11
100	27.21	1.05
200	27.43	0.25
300	27.54	0.13
500	27.49	0.03

Moreover, Assumption 3.5 is satisfied since the process increases but Assumption 3.6 is not, because $t^*(x) = +\infty$ for all $x \in E$. However, as pointed out in Section 3, this can be solved by considering the process killed at time τ .

The mean exit time. Table 1 displays the simulation results for the approximation of the mean exit time. For different numbers of points in the quantization grids, the value of $\widehat{p}_{N,1}$ which approximates the mean exit time is given. A reference value is obtained thanks to the Monte Carlo method (10^6 simulations): $E[\tau_{10}]_{\text{Monte Carlo}} = 5.125$.

The second moment. We present the results of the approximation of the second moment in Table 2. Our Monte Carlo reference value (10^6 simulations) is $E[\tau_{10}^2]_{\text{Monte Carlo}} = 27.5$.

For the first and second moments, the empirical convergence rates are presented in Figure 2. Through a regression model the empirical convergence is estimated as -1.23 for the first moment and -1.39 for the second moment. Note that they are roughly of the same order as the rate of convergence of the optimal quantizer (see Theorem 4.1), as here the dimension is 1.

The exit time distribution. As mentioned earlier, we can obtain the exact value of the survival function of the exit time.

Proposition 5.1. *Denote by $\text{fl}(\cdot)$ the floor function. For all $s, b \in \mathbb{R}^+$, we have*

$$P(\tau_b \geq s) = \begin{cases} P(T_{\text{fl}(b-s)+1} > s) & \text{for all } s \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

Remark 5.1. Note that T_k has distribution $\gamma(k, 1)$, so the right-hand side term in the above proposition can be computed easily.

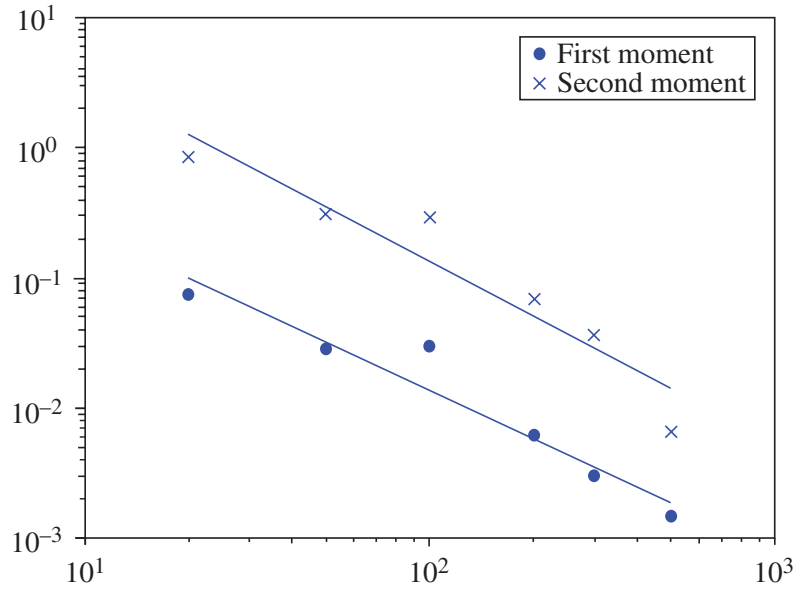


FIGURE 2: Logarithm of the error with respect to the logarithm of the number of points in the quantization grids for the first and second moments of the Poisson process.

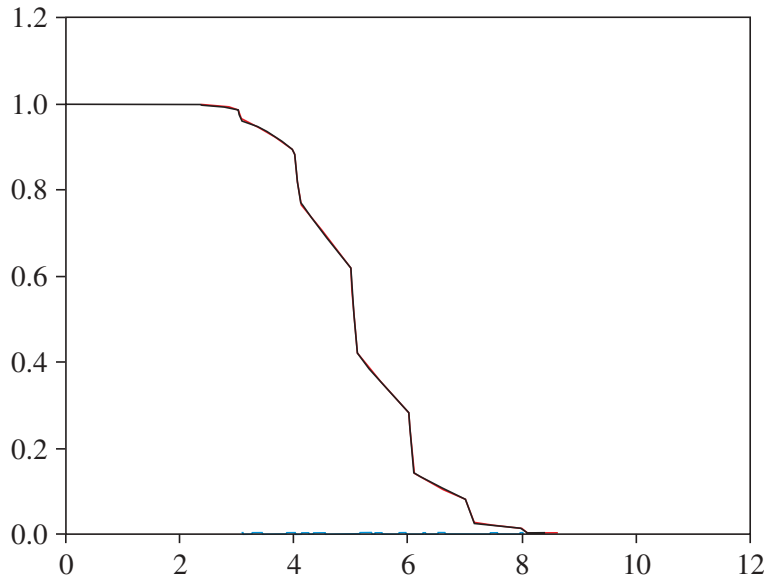


FIGURE 3: Survival function of τ_{10} and its quantized approximation with 500 points in the quantization grids. The functions appear indistinguishable.

Proof of Proposition 5.1. Let $s > 0$. Note that $Y_s \geq s$; thus, $\tau_b < s$ a.s. when $s > b$. Assume now that $s \leq b$. We have

$$P(\tau_b \geq s) = P(Y_s \leq b) = P(N_s \leq b - s) = P(N_s \leq \text{fl}(b - s)) = P(T_{\text{fl}(b-s)+1} \geq s).$$

This completes the proof.

Figure 3 shows both the exact survival function of the exit time and its quantized approximation. Table 3 contains the empirical error between the two functions. For the survival function, the empirical convergence rate is presented in Figure 4. Through a regression model the convergence rate is estimated as -1.05 . Note that it is roughly of the same order as the rate of convergence of the optimal quantizer (see Theorem 4.1), as here the dimension is 1.

TABLE 3: Simulation results for the distribution.

Number of points in the quantization grids	$\max_s p_N(s) - \hat{p}_N(s) $
20	0.090
50	0.077
100	0.057
200	0.011
300	0.007
500	0.005

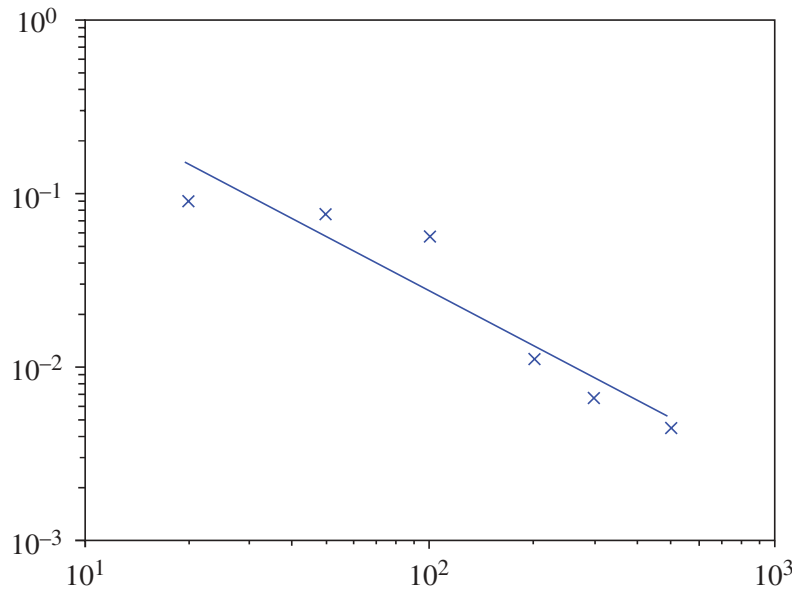


FIGURE 4: Logarithm of the error with respect to the logarithm of the number of points in the quantization grids for the survival function of the Poisson process.

Remark 5.2. We already insisted on the fact that our approach is flexible with respect to U . In this example we could very quickly obtain the mean exit time or the exit time distribution for a different set $U' = (-\infty, b']$ for any $0 < b' \leq b = 10$. Indeed, $P(\tau_{b'} > T_{10}) = 0$, so it is not necessary to compute new quantization grids.

Remark 5.3. Recall that the value of T_k may be obtained from Z_k since $T_k = Z_k - k$, so it is sufficient to quantize the process $(Z_k)_{k \leq N}$ instead of $(Z_k, T_k)_{k \leq N}$. The reduction of the dimension of the process that has to be quantized results in an improvement of the convergence rate and it appears that the approximations presented in the previous tables indeed converge very quickly.

Convergence rate for the exit time distribution. We note from the proof of Proposition 4.2 that a bound for the rate of convergence for the exit time distribution can be obtained as soon as, for all $k \in \{0, \dots, N-1\}$, the survival function of $(T_k + u^*(Z_k)) \wedge T_{k+1}$ denoted φ_k is piecewise Lipschitz continuous. Although it is difficult to state general assumptions under which this is true, the following proposition proves that the condition is fulfilled in our example.

Proposition 5.2. *For all $k \in \{0, \dots, N-1\}$, the survival function φ_k of $(T_k + u^*(Z_k)) \wedge T_{k+1}$ is Lipschitz continuous on $(-\infty; b-k)$ and on $(b-k; +\infty)$ with Lipschitz constant $[\varphi_k] \leq 1$.*

Proof. Let $k = 0$ and $s > 0$. We have

$$\begin{aligned}\varphi_0(s) &= \mathbf{P}_\mu((T_0 + u^*(Z_0)) \wedge T_1 > s) \\ &= \mathbf{P}_\mu(b \wedge T_1 > s) \\ &= \mathbf{1}_{\{b>s\}} \mathbf{P}_\mu(T_1 > s) \\ &= \mathbf{1}_{\{b>s\}} e^{-s} \quad (\text{since } T_1 \text{ has exponential distribution with parameter } 1).\end{aligned}$$

Therefore, the function φ_0 is equal to 0 on $[b; +\infty)$ and is Lipschitz continuous with Lipschitz constant 1 on $(0; b)$.

Let $k \geq 1$ and $s > 0$, and recall that the random variables $(S_j)_{j \geq 0}$ are independent and have exponential distributions with parameter 1 so that, in particular, T_k and S_{k+1} are independent and T_k has distribution $\gamma(k, 1)$. Moreover, recall that $Z_k = k + T_k$ and that $u^*(x) = b - x$. Then

$$\begin{aligned}\varphi_k(s) &= \mathbf{P}_\mu((T_k + u^*(Z_k)) \wedge T_{k+1} > s) \\ &= \int_{(\mathbb{R}^+)^2} \mathbf{1}_{\{(t+(b-k-t)) \wedge u > s\}} f_{\gamma(k,1)}(t) f_{\gamma(k+1,1)}(u) dt du,\end{aligned}$$

where $f_{\gamma(j,1)}$ denotes the density function of the distribution $\gamma(j, 1)$ for $j \in \{k, k+1\}$.

Let $s' > s > 0$. We have

$$\begin{aligned}|\varphi_k(s') - \varphi_k(s)| &\leq \int_{(\mathbb{R}^+)^2} |\mathbf{1}_{\{(b-k) \wedge u > s'\}} - \mathbf{1}_{\{(b-k) \wedge u > s\}}| f_{\gamma(k,1)}(t) f_{\gamma(k+1,1)}(u) dt du \\ &\leq \int_{(\mathbb{R}^+)^2} \mathbf{1}_{\{(b-k) \wedge u \in (s; s']\}} f_{\gamma(k,1)}(t) f_{\gamma(k+1,1)}(u) dt du \\ &\leq \int_{(\mathbb{R}^+)^2} (\mathbf{1}_{\{b-k \in (s; s']\}} + \mathbf{1}_{\{u \in (s; s']\}}) f_{\gamma(k,1)}(t) f_{\gamma(k+1,1)}(u) dt du \\ &\leq \mathbf{1}_{\{b-k \in (s; s']\}} + C_{f_{\gamma(k+1,1)}} |s' - s| \\ &\leq \mathbf{1}_{\{b-k \in (s; s']\}} + |s' - s| \quad \left(\text{since } C_{f_{\gamma(k+1,1)}} = \frac{1}{(k)!} \left(\frac{k}{e}\right)^k \leq 1 \right).\end{aligned}$$

If s and s' both belong to $(0; b - k)$ or if they both belong to $(b - k; +\infty)$, we have $|\varphi_k(s') - \varphi_k(s)| \leq |s' - s|$. This completes the proof.

Consequently, in this example, we are now able to state a bound for the rate of convergence of the exit time distribution approximation scheme. The following proposition is therefore an improvement over Proposition 4.2 and Theorem 4.2.

Proposition 5.3. *For all $k \in \{0, \dots, N-1\}$, let $s > 0$ and assume that the quantization error is small enough to ensure that*

$$\left(\frac{p}{2}\right)^{1/(p+1)} (\|T_k - \widehat{T}_k\|_p + \|Z_k - \widehat{Z}_k\|_p + \|T_{k+1} - \widehat{T}_{k+1}\|_p)^{p/(p+1)} < |b - k - s|.$$

Then we have

$$\begin{aligned}|r_k(s) - \widehat{r}_k(s)| &\leq 2 \left(\frac{p}{2}\right)^{1/(p+1)} \left(\frac{1}{p} + 1\right) \\ &\quad \times (\|T_k - \widehat{T}_k\|_p + \|Z_k - \widehat{Z}_k\|_p + \|T_{k+1} - \widehat{T}_{k+1}\|_p)^{p/(p+1)} \\ &\quad + |q_k - \widehat{q}_k| + |q_{k+1} - \widehat{q}_{k+1}|.\end{aligned}$$

Moreover, for all $k \in \{0, \dots, N\}$, if the quantization error is such that, for all $l \leq k$,

$$2\left(\frac{p}{2}\right)^{1/(p+1)}\left(\frac{1}{p} + 1\right)\|Z_l - \widehat{Z}_l\|_p^{p/(p+1)} \leq \frac{1}{2}\tilde{q},$$

then we have

$$\begin{aligned} |p_k(s) - \widehat{p}_k(s)| &\leq \frac{2}{\tilde{q}}(|q_{k-1} - \widehat{q}_{k-1}| + |p_{k-1}(s) - \widehat{p}_{k-1}(s)| + |r_{k-1}(s) - \widehat{r}_{k-1}(s)|) \\ &\quad + \frac{4}{\tilde{q}^2}|q_k - \widehat{q}_k|. \end{aligned}$$

Proof. The proof follows directly from the proofs of Proposition 4.2 and Theorem 4.2. Simply note that the A term may be bounded thanks to the piecewise Lipschitz continuity of the functions φ_k on $(-\infty; b - k)$ and on $(b - k; +\infty)$. Let $s > 0$, $s \neq b - k$, and let $\alpha > 0$ such that $b - k \notin [s - \alpha; s + \alpha]$, i.e. $\alpha < |b - k - s|$. Then

$$\begin{aligned} A &\leq |\varphi_k(s + \alpha) - \varphi_k(s - \alpha)| + \frac{\|\eta\|_p^p}{\alpha^p} \quad (\text{from the proof of Proposition 4.2}) \\ &\leq 2[\varphi_k]\alpha + \frac{\|\eta\|_p^p}{\alpha^p}, \end{aligned}$$

which reaches a minimum when $\alpha = (p\|\eta\|_p^p/2[\varphi_k])^{1/(p+1)}$. Note that $[\varphi_k] = 1$ and $[u^*] = 1$.

Remark 5.4. We can calculate the exact value of \tilde{q} that is the first nonnegative value of the sequence $(\mathbb{P}_\mu(Z_k \notin U))_k$. We have $\tilde{q} = \mathbb{P}_\mu(Z_1 \notin (-\infty; 10)) = \mathbb{P}_\mu(T_1 \geq 9) = e^{-9}$ because T_1 has an exponential distribution with parameter 1.

5.2. A corrosion model example

Let us consider the structure of aluminium corroded successively in three different environments. Corrosion is prevented by some protection until a random time γ when corrosion starts. Then, in each environment $i \in \{1; 2; 3\}$, the loss of thickness satisfies

$$d_i(t) = \rho_i(t - \gamma + \eta_i(e^{-(t-\gamma)/\eta_i} - 1)) \mathbf{1}_{\{t \geq \gamma\}},$$

where ρ_i is the corrosion rate (ρ_i has a uniform distribution on an interval that depends on the environment i) and η_i is a constant transition time. The structure goes from environment 1 to environment 2, then from 2 to 3, from 3 to 1, and so on. It remains in environment i for a time T_i , which has an exponential distribution with parameter λ_i . When the loss of thickness reaches 0.2 mm, the piece is said to be unusable; this will be the exit criterion. Table 4 gives the values of the different parameters.

The loss of thickness will be represented by a PDMP whose modes are the different environments. Let $M = \{(i, j): i \in \{1, 2, 3\}, j \in \{0, 1\}\}$. For $m = (i, j) \in M$, i represents the environment and j is worth 1 if the protection γ is still active and 0 otherwise. For each $m \in M$, let $E_m = \mathbb{R}^4$ and, for $\xi \in E_m$, ξ represents the family (d, s, ρ, γ) , where d is the corroded thickness and s is the time since the last jump. The set U_m will therefore be, for all $m \in M$, $U_m = (-\infty; 0.2] \times \mathbb{R}^3$. This set is convex, so Assumption 3.3 is satisfied. Finally, the

TABLE 4: Numerical values of the parameters of the corrosion model.

	Environment		
	1	2	3
λ_i (h^{-1})	$(17\,520)^{-1}$	$(131\,400)^{-1}$	$(8760)^{-1}$
η_i (h)	30 000	200 000	40 000
ρ_i (mm/h)	$[10^{-6}, 10^{-5}]$	$[10^{-7}, 10^{-6}]$	$[10^{-6}, 10^{-5}]$
γ (h)	Weibull distribution with $\alpha = 2.5$ and $\beta = 11\,800$		

flow in mode $m = (i, j)$ is

$$\Phi_{(i,0)} \left(\begin{pmatrix} d \\ s \\ \rho \\ 0 \end{pmatrix}, t \right) = \begin{pmatrix} d + d_m(t+s) - d_m(s) \\ t+s \\ \rho \\ 0 \end{pmatrix},$$

$$\Phi_{(i,1)} \left(\begin{pmatrix} 0 \\ s \\ \rho \\ \gamma \end{pmatrix}, t \right) = \begin{pmatrix} 0 \\ t+s \\ \rho \\ (\gamma - t) \mathbf{1}_{\{\gamma \geq t\}} \end{pmatrix}.$$

The parameters d and γ evolve continuously between the jumps, but ρ is chosen independently after each jump and is constant along the flow.

Let us consider the approximation of the distribution and of the mean exit time. Consider the first moment. We note that $E_\mu[\tau] = E_\mu[\gamma] + E_\mu[\tau']$, where γ has Weibull distribution and τ' represents the exit time in the case of a process without initial protection against corrosion (i.e. $\gamma = 0$). Therefore, it is sufficient to check whether τ' satisfies the required assumptions. Hence, let $\gamma = 0$ and note that u^* is then bounded since $\rho \geq 10^{-7}$ and $\eta \leq 200\,000$, so $d_m(t) \geq 10^{-7}(t - 200\,000)$ and eventually $u^* \leq 0.2 \times 10^7 + 200\,000 = 2.2 \times 10^6$ h. Denote this bound by C_{u^*} . Consider the distribution. Assumption 3.2(b) (the boundedness condition on u^*) is not required according to Remark 4.3. Moreover, from the proofs of Propositions 4.2 and 4.4, it follows that Assumption 3.2(a) (the Lipschitz continuity condition on u^*) becomes useless in this example thanks to Lemma 5.3. Assumption 3.4 follows from Lemma 5.2 below. Eventually, Assumption 3.5 is satisfied, but Assumption 3.6 is not. However, considering the process killed at time τ solves this issue.

Lemma 5.2. *For all $\alpha > 0$ and all $k \in \{0, \dots, N\}$,*

$$P_\mu(Z_k \in U^\alpha) \leq 5\alpha.$$

Proof. For notational convenience, let M_k, D_k, R_k , and G_k denote the values of m, d, ρ , and γ after the k th jump, so $Z_k = (M_k, D_k, R_k, G_k)$. Note now that

$$P_\mu(Z_k \in U^\alpha) = P_\mu(|D_k - 0.2| \leq \alpha).$$

We therefore study more precisely the law of D_k . Let $K = \inf\{k \geq 0 \text{ such that } G_k = 0\}$; K is the jump that occurs at the end of the protection period against corrosion. Define $F(s) = s + \eta(e^{-s/\eta} - 1)$. Then we have

$$D_k = \begin{cases} 0 & \text{for } k \leq K, \\ D_{k-1} + R_k F(S_k) & \text{for } k > K. \end{cases}$$

Let us now prove that, for all k , the random variable $R_k F(S_k)$ has a bounded density. Recall that R_k has a uniform distribution on $[a_k; b_k] \subset [10^{-7}; 10^{-5}]$ and that S_k has an exponential distribution with parameter λ_k . Now let h be a real, bounded, measurable function. Then

$$\mathbb{E}_\mu[h(R_k F(S_k))] = \int_0^{+\infty} \int_{a_k}^{b_k} h(\rho F(s)) \frac{1}{b_k - a_k} \lambda_k e^{-\lambda_k s} d\rho ds.$$

Introduce the transformation

$$u = \rho, \quad v = \rho F(s),$$

whose Jacobian is worth $(1/u)(F^{-1})'(v/u)$, so

$$\mathbb{E}_\mu[h(R_k F(S_k))] = \int_0^{+\infty} h(v) \left(\int_{a_k}^{b_k} \frac{\lambda_k e^{-\lambda_k F^{-1}(v/u)} (F^{-1})'(v/u)}{(b_k - a_k)u} du \right) dv.$$

Hence, we obtain the density of the random variable $R_k F(S_k)$ and integration by parts yields

$$\begin{aligned} \int_{a_k}^{b_k} \frac{\lambda_k e^{-\lambda_k F^{-1}(v/u)} (F^{-1})'(v/u)}{(b_k - a_k)u} du &= \frac{1}{b_k - a_k} \int_{a_k}^{b_k} \frac{\lambda_k e^{-\lambda_k F^{-1}(v/u)} (F^{-1})'(v/u)}{u^2} du \\ &= \frac{1}{b_k - a_k} \left([u e^{-\lambda_k F^{-1}(v/u)}]_{a_k}^{b_k} - \int_{a_k}^{b_k} e^{-\lambda_k F^{-1}(v/u)} du \right). \end{aligned}$$

Finally, the density of the random variable $R_k F(S_k)$ is bounded by

$$\left| \int_{a_k}^{b_k} \frac{\lambda_k e^{-\lambda_k F^{-1}(v/u)} (F^{-1})'(v/u)}{(b_k - a_k)u} du \right| \leq \frac{a_k + b_k}{b_k - a_k} + 1 \leq \frac{2b_k}{b_k - a_k} \leq 2.$$

Let $j \in \mathbb{N}$. We now study the distribution of the random variables $(D_k)_{k \in \mathbb{N}}$ conditionally on the event $\{K = j\}$. An induction argument shows that, conditionally on the event $\{K = j\}$, the random variable D_k has distribution δ_0 for $k \leq j$ and has a density ψ_k bounded by 2 for $k > j$. Indeed, in the second case, the density of D_k may be obtained by convolution since D_{k-1} and $R_k F(S_k)$ are independent random variables. Therefore, for $k \leq j$,

$$\mathbb{P}_\mu(|D_k - 0.2| \leq \alpha \mid K = j) = \mathbf{1}_{\{\alpha \geq 0.2\}} \leq 5\alpha$$

since $D_k = 0$ for $k \leq j$ and, for $k > j$,

$$\mathbb{P}_\mu(|D_k - 0.2| \leq \alpha \mid K = j) = \int_{0.2-\alpha}^{0.2+\alpha} \psi_k(v) dv \leq 4\alpha$$

since $\psi_k \leq 2$. Eventually,

$$\mathbb{P}_\mu(Z_k \in U^\alpha) = \mathbb{P}_\mu(|D_k - 0.2| \leq \alpha) = \sum_{j \in \mathbb{N}} \mathbb{P}_\mu(|D_k - 0.2| \leq \alpha \mid K = j) \mathbb{P}_\mu(K = j) \leq 5\alpha.$$

This completes the proof.

Lemma 5.3. For all $k \in \mathbb{N}$, let

$$\eta_k = |((T_k + u^*(Z_k)) \wedge T_{k+1}) - ((\widehat{T}_k + u^*(\widehat{Z}_k)) \wedge \widehat{T}_{k+1})|.$$

We have, for all $\alpha > 0$,

$$\|\eta_k\|_p \leq \|T_k - \widehat{T}_k\|_p + 2\|T_{k+1} - \widehat{T}_{k+1}\|_p + \left([u^*]_{\alpha/2} + \frac{4C_{u^*}}{\alpha}\right)\|Z_k - \widehat{Z}_k\|_p + 10C_{u^*}\alpha^{1/p},$$

where $[u^*]_{\alpha} = (1 + C_{u^*} + 4 \times 10^5)/(10^{-7}(1 - e^{-\alpha/2}))$.

Proof. Let $\alpha > 0$. Let $\tilde{U}_{\alpha} = [0, 0.2 - \alpha] \times \{0\} \times [10^{-7}; 10^{-5}] \times \{0\}$. We will prove that the function $u^*(d, 0, \rho, 0)$ is Lipschitz continuous on this set. The function $u^*(d, 0, \rho, 0)$ satisfies the following equivalent equations:

$$d + d_m(u^*) = 0.2 \iff d + \rho(u^* + \eta(e^{-u^*/\eta} - 1)) = 0.2.$$

The implicit equation satisfied by u^* yields, on the set \tilde{U}_{α} , $u^* \geq \alpha/\rho_{\max} = 10^5\alpha$. This lower bound will be crucial to prove the Lipschitz continuity. Let $d, d' \leq 0.2 - \alpha$, and define $u = u^*(d, 0, \rho, 0)$ and $u' = u^*(d', 0, \rho, 0)$. Note that $d + d_m(u) = d' + d_m(u')$ because they are both equal to 0.2. Consequently, $|d_m(u) - d_m(u')| = |d' - d|$ and, noting that $\eta \leq 2 \times 10^5$, we have

$$\begin{aligned} |d - d'| &= \rho|u - u' + \eta(e^{-u/\eta} - e^{-u'/\eta})| \\ &\geq \rho(1 - e^{-u \wedge u'/\eta})|u - u'| \\ &\geq 10^{-7}(1 - e^{-\alpha/2})|u - u'|, \end{aligned}$$

which proves the Lipschitz continuity of u^* with respect to d on \tilde{U}_{α} .

Similarly, let $\rho, \rho' \in [10^{-7}; 10^{-5}]$, and define $u = u^*(d, 0, \rho, 0)$ and $u' = u^*(d, 0, \rho', 0)$. Note that $d + \rho(u + \eta(e^{-u/\eta} - 1)) = d + \rho'(u' + \eta(e^{-u'/\eta} - 1))$ because they are both equal to 0.2. Subtracting $d + \rho(u' + \eta(e^{-u'/\eta} - 1))$ from both terms yields

$$\rho|u - u' + \eta(e^{-u/\eta} - e^{-u'/\eta})| = |\rho - \rho'| |u' + \eta(e^{-u'/\eta} - 1)|.$$

A lower bound for the left-hand side term has already been computed, while the right hand-side is easily bounded by $(C_{u^*} + 4 \times 10^5)|\rho - \rho'|$, since $\eta \leq 2 \times 10^5$, so we have

$$(C_{u^*} + 4 \times 10^5)|\rho - \rho'| \geq 10^{-7}(1 - e^{-\alpha/2})|u - u'|,$$

which proves the Lipschitz continuity of u^* with respect to ρ on \tilde{U}_{α} . Eventually, for all $\alpha > 0$, the function u^* is Lipschitz continuous on \tilde{U}_{α} with Lipschitz constant $[u^*]_{\alpha} = (1 + C_{u^*} + 4 \times 10^5)/10^{-7}(1 - e^{-\alpha/2})$.

Let $k \in \mathbb{N}$. We now intend to bound $\|\eta_k\|_p$. Define, as in the proof of Lemma 5.2, the random variable $K = \inf\{k \geq 0 \text{ such that } G_k = 0\}$; K is the jump that occurs at the end of the protection period against corrosion.

First, note that, on the event $\{k \leq K\}$ (i.e. when protection against corrosion is still active), we have $Z_k \in E_{(i,1)}$ for some $i \in \{1, 2, 3\}$ and, since the projection defining \widehat{Z}_k from Z_k ensures that they are in the same mode, we also have $\widehat{Z}_k \in E_{(i,1)}$. Moreover, $u^*(x) = +\infty$ for all $x \in E_{(i,1)}$, so

$$\|\eta_k \mathbf{1}_{\{k \leq K\}}\|_p = \|(T_{k+1} - \widehat{T}_{k+1}) \mathbf{1}_{\{k \leq K\}}\|_p \leq \|T_{k+1} - \widehat{T}_{k+1}\|_p.$$

Furthermore, if $Z_k = \Delta$, where Δ denotes the cemetery state, then $\widehat{Z}_k = \text{proj}_{\Gamma_k}(Z_k) = \Delta$ too and we have $\eta_k = 0$, so

$$\begin{aligned} \|\eta_k \mathbf{1}_{\{k > K\}}\|_p &\leq \|\eta_k \mathbf{1}_{\{k > K\}} \mathbf{1}_{\{Z_k \neq \Delta\}}\|_p \\ &\leq \|T_k - \widehat{T}_k\|_p + \|T_{k+1} - \widehat{T}_{k+1}\|_p + \|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{k > K\}} \mathbf{1}_{\{Z_k \neq \Delta\}}\|_p. \end{aligned}$$

Eventually, we intend to bound the last term of the previous sum; therefore, we consider the event $\{k > K\} \cap \{Z_k \neq \Delta\}$. On the one hand, the random variables Z_k and \widehat{Z}_k both belong to $E_{(i,0)}$ for some $i \in \{1, 2, 3\}$. On the other hand, although $U_m = (-\infty; 0.2] \times \mathbb{R}^3$ for all $m \in M$, we actually have $Z_k \in [0; 0.2] \times \{0\} \times [10^{-7}; 10^{-5}] \times \mathbb{R}^+$ a.s. and, according to Remark 4.1, $\widehat{Z}_k \in [0; 0.2] \times \{0\} \times [10^{-7}; 10^{-5}] \times \mathbb{R}^+$ a.s. too. Combining the two previous remarks, we have $Z_k \in \tilde{U}$ and $\widehat{Z}_k \in \tilde{U}$, where $\tilde{U} = [0; 0.2] \times \{0\} \times [10^{-7}; 10^{-5}] \times \{0\}$. Finally, let $\alpha > 0$ and note that $\tilde{U} \subset \tilde{U}_\alpha \cup U^\alpha$. We have

$$\|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{k \geq K\}} \mathbf{1}_{\{Z_k \neq \Delta\}}\|_p \leq A + B,$$

where

$$A = \|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{Z_k \in \tilde{U}_\alpha\}} \mathbf{1}_{\{k \geq K\}}\|_p, \quad B = \|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{Z_k \in U^\alpha\}} \mathbf{1}_{\{k \geq K\}}\|_p.$$

The term B is easily bounded thanks to Lemma 5.2: $B \leq 2C_{u^*} \mathbb{P}_\mu(Z_k \in U^\alpha)^{1/p} \leq 10C_{u^*} \alpha^{1/p}$. We now turn to the term A and use the Lipschitz continuity of u^* on \tilde{U}_β for any $\beta > 0$. We have

$$\begin{aligned} A &\leq \|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{Z_k \in \tilde{U}_\alpha\}} \mathbf{1}_{\{\widehat{Z}_k \in \tilde{U}_{\alpha/2}\}} \mathbf{1}_{\{k \geq K\}}\|_p \\ &\quad + \|(u^*(Z_k) - u^*(\widehat{Z}_k)) \mathbf{1}_{\{Z_k \in \tilde{U}_\alpha\}} \mathbf{1}_{\{\widehat{Z}_k \notin \tilde{U}_{\alpha/2}\}} \mathbf{1}_{\{k \geq K\}}\|_p \\ &\leq [u^*]_{\alpha/2} \|Z_k - \widehat{Z}_k\|_p + 2C_{u^*} \|\mathbf{1}_{\{Z_k \in \tilde{U}_\alpha\}} \mathbf{1}_{\{\widehat{Z}_k \notin \tilde{U}_{\alpha/2}\}}\|_p. \end{aligned}$$

Note now that $\mathbf{1}_{\{Z_k \in \tilde{U}_\alpha\}} \mathbf{1}_{\{\widehat{Z}_k \notin \tilde{U}_{\alpha/2}\}} \leq \mathbf{1}_{\{|Z_k - \widehat{Z}_k| \geq \alpha/2\}}$, so, finally,

$$\begin{aligned} A &\leq [u^*]_{\alpha/2} \|Z_k - \widehat{Z}_k\|_p + 2C_{u^*} \left(\mathbb{P}_\mu \left(|Z_k - \widehat{Z}_k| \geq \frac{\alpha}{2} \right) \right)^{1/p} \\ &\leq [u^*]_{\alpha/2} \|Z_k - \widehat{Z}_k\|_p + 4C_{u^*} \frac{\|Z_k - \widehat{Z}_k\|_p}{\alpha}, \end{aligned}$$

completing the proof.

The mean exit time. Simulation results for the approximation of the mean exit time are given in Table 5. In order to have a value of reference, a Monte Carlo method (10^6 simulations) yields the value $\mathbb{E}[\tau]_{\text{Monte Carlo}} = 526 \times 10^3$ h. For the first moment, the empirical convergence rate is presented in Figure 5. Through a regression model the empirical convergence rate is estimated as -0.38 . Note that it is roughly of the same order as the rate of convergence of the optimal quantizer (see Theorem 4.1), as here the dimension is 4.

TABLE 5: Simulation results for the mean exit time.

Number of points in the quantization grids	$\widehat{p}_{N,1} (\times 10^3 \text{ h})$	Relative error to $526 \times 10^3 \text{ h}$ (%)
20	572	8.7
50	569	8.2
100	557	5.9
200	551	4.8
500	539	2.5

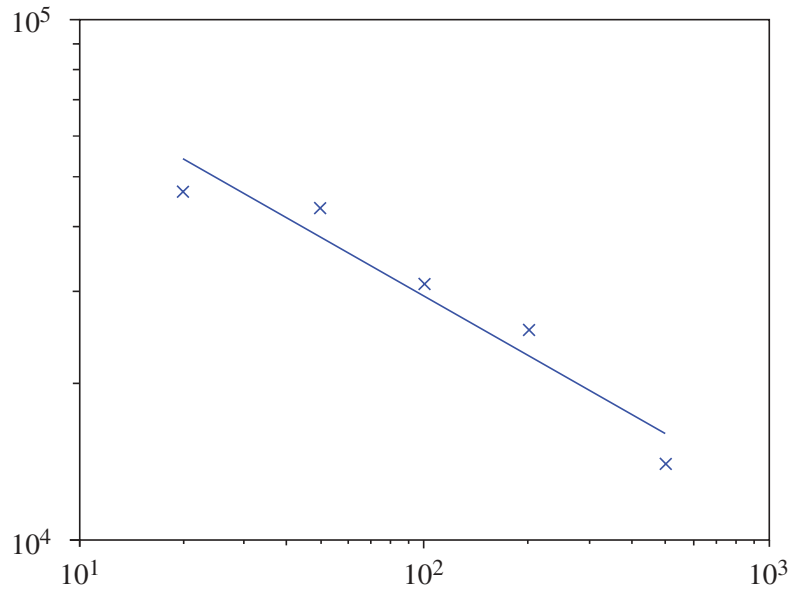


FIGURE 5: Logarithm of the error with respect to the logarithm of the number of points in the quantization grids for the first moment of the corrosion process.

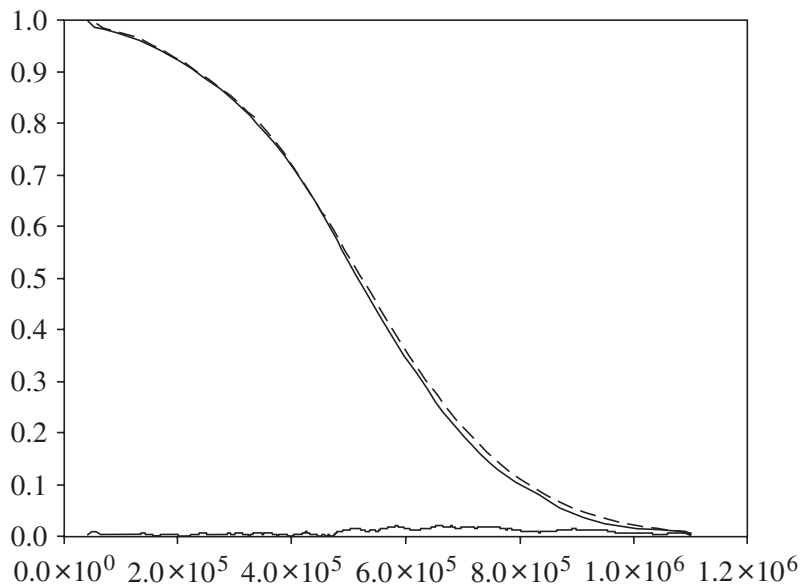


FIGURE 6: Survival function of τ obtained using Monte Carlo simulations (*dashed line*) and the quantized approximation (*solid line*), and the error with 500 points in the quantization grids.

The exit time distribution. Considering the approximation scheme for the exit time distribution, we note that the quantized value $\hat{p}_N(s)$ is not necessarily smaller than 1. Therefore, it appears natural to replace $\hat{p}_N(s)$ by $\hat{p}_N(s) \wedge 1$. This does not change the convergence theorem and can only improve the approximation error. It is equally possible, and this is done in the results below, to replace $\hat{p}_N(s)$ by $\hat{p}_N(s)/\hat{p}_N(0)$ since $\hat{p}_N(0)$ goes to 1.

Figure 6 presents the survival function of τ obtained using Monte Carlo simulations (dashed line) and our approximation scheme (solid line), and the error. Table 6 contains the empirical error for different numbers of points in the quantization grids. For the survival function, the empirical convergence rate is presented in Figure 7. Through a regression model the empirical

TABLE 6: Simulation results for the distribution.

Number of points in the quantization grids	$\max_s p_N(s) - \hat{p}_N(s) $
20	0.145
50	0.119
100	0.040
200	0.039
500	0.020

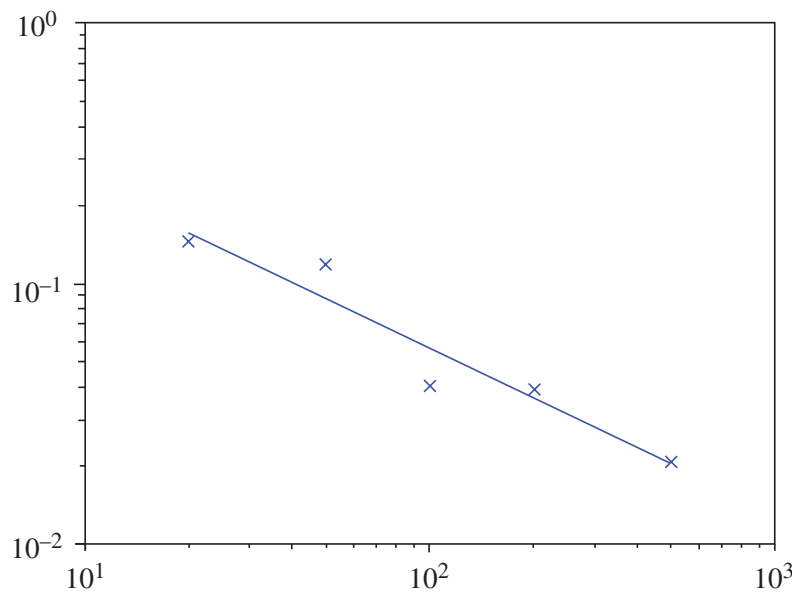


FIGURE 7: Logarithm of the error with respect to the logarithm of the number of points in the quantization grids for the survival of the corrosion process.

convergence rate is estimated as -0.63 . Note that it is roughly of the same order as the rate of convergence of the optimal quantizer (see Theorem 4.1), as here the dimension is 4.

The convergence of the approximation scheme in the corrosion model appears to be slightly slower than in the previous example. This is due to the higher dimension of the process that has to be quantized, which is 4 in the case of the corrosion model and 1 in the case of the Poisson process.

Remark 5.5. Using Monte Carlo simulations, we can approximate the value of \tilde{q} . We have $\tilde{q} \simeq 0.0187$ for 10^7 histories.

6. Advantages and practical interest of our approach

Let us describe the practical interest of our approach.

- The quantization grids only have to be computed once and can be used for several purposes. Moreover, once they are obtained, the procedures leading to $\hat{p}_N(s)$ and to $\hat{p}_{N,j}$ can be achieved very simply since we only have to compute finite sums.
- Concerning the distribution, since $\hat{p}_N(s)$ can be computed almost instantly for any value of s , the whole survival function can be obtained very quickly. Similarly, concerning

the moments, $\widehat{p}_{N,j}$ can be computed very quickly for any j , so any moment is almost instantly available.

- Furthermore, in both cases, one may decide to change the set U and consider the exit time τ' from a new set U' . This will yield new sequences $(\widehat{q}_k)_k$, $(\widehat{r}_{k,j})_k$, and $(\widehat{p}_{k,j})_k$ in the case of the j th moment approximation or new sequences $(\widehat{q}_k)_k$, $(\widehat{r}_k(s))_k$, and $(\widehat{p}_k(s))_k$ if we are interested in the distribution. These new sequences are obtained quickly and easily since the quantized process remains the same and we only have to compute finite sums. Of course, the set U' must be such that Assumptions 3.2–3.5 remain true and such that $P_\mu(T_N < \tau')$ remains small without changing the computation horizon N . This last condition is fulfilled if, for instance, $U' \subset U$. This flexibility is an important advantage of our method over, for instance, a Monte Carlo method.

Acknowledgements

This work was supported by the ARPEGE program of the French National Agency of Research (ANR), project ‘FAUTOCOES’, grant number ANR-09-SEGI-004. The authors also gratefully acknowledge EADS Astrium for its financial support.

References

- [1] BALLY, V. AND PAGÈS, G. (2003). A quantization algorithm for solving multi-dimensional discrete-time optimal stopping problems. *Bernoulli* **9**, 1003–1049.
- [2] BALLY, V., PAGÈS, G. AND PRINTEMPS, J. (2005). A quantization tree method for pricing and hedging multidimensional American options. *Math. Finance* **15**, 119–168.
- [3] BOUTON, C. AND PAGÈS, G. (1997). About the multidimensional competitive learning vector quantization algorithm with constant gain. *Ann. Appl. Prob.* **7**, 679–710.
- [4] CHIQUET, J. AND LIMNIOS, N. (2008). A method to compute the transition function of a piecewise deterministic Markov process with application to reliability. *Statist. Prob. Lett.* **78**, 1397–1403.
- [5] DAVIS, M. H. A. (1993). *Markov Models and Optimization* (Monogr. Statist. Appl. Prob. **49**). Chapman & Hall, London.
- [6] DE SAPORTA, B., DUFOUR, F. AND GONZALEZ, K. (2010). Numerical method for optimal stopping of piecewise deterministic Markov processes. *Ann. Appl. Prob.* **20**, 1607–1637.
- [7] GRAY, R. M. AND NEUHOFF, D. L. (1998). Quantization. *IEEE Trans. Inf. Theory* **44**, 2325–2383.
- [8] HELMES, K., RÖHL, S. AND STOCKBRIDGE, R. H. (2001). Computing moments of the exit time distribution for Markov processes by linear programming. *Operat. Res.* **49**, 516–530.
- [9] LASSERRE, J.-B. AND PRIETO-RUMEAU, T. (2004). SDP vs. LP relaxations for the moment approach in some performance evaluation problems. *Stoch. Models* **20**, 439–456.
- [10] PAGÈS, G., PHAM, H. AND PRINTEMPS, J. (2004). Optimal quantization methods and applications to numerical problems in finance. In *Handbook of Computational and Numerical Methods in Finance*, Birkhäuser, Boston, MA, pp. 253–297.

4 Articles relatifs au Chapitre 4

- 4.1 [NML⁺12] A. NÈGRE, O. MARCEAU, D. LANEUVILLE, H. ZHANG, B. DE SAPORTA, AND F. DUFOUR. Stochastic control for underwater optimal trajectories. In *IEEE Aerospace conference*, Big Sky, Montana, USA, 2012.

Stochastic Control for Underwater Optimal Trajectories

Adrien Nègre
DCNS Research
Toulon, France
adrien.negre@dcnsgroup.com

Olivier Marceau
DCNS Research
Paris, France
olivier.marceau@dcnsgroup.com

Dann Laneuville
DCNS Research
Paris, France
dann.laneuville@dcnsgroup.com

Huilong Zhang
CQFD team, INRIA
Bordeaux, France
huilong.zhang@math.u-bordeaux1.fr

Benoite De Saporta
CQFD team, INRIA
Bordeaux, France
benoite.de-saporta@u-bordeaux4.fr

Francois Dufour
CQFD team, INRIA
Bordeaux, France
dufour@math.u-bordeaux1.fr

Abstract—This paper¹ describes an original approach based on dynamic programming theory (discrete-time finite horizon Markov control) to address the difficult problem of computing optimal trajectories with respect to some criteria for any vehicle evolving in a given environment (in a probabilistic point of view) to accomplish some tasks (defined in a complex Lipschitz criteria). After a brief remind about dynamic programming The whole control process is detailed and applied in this paper for a submarine equipped with a sonar system which wants to detect targets as well as possible. Promising result about the optimal depth control for a submarine are presented with one and then two targets showing that the chosen approach is a good candidate.

well known curse of dimensionality. Hence, solving MDPs numerically is an awkward and important problem.

The method is applied to control a submarine which wants to well detect one or several targets. Why? A smart operator, if provided information about target's position and velocity and a sound propagation code can find a good trajectory. If we now consider a submarine surrounded by several targets, it is clear that a human operator will have great difficulty to find the best route.

Section 2 states the problem in a general context (required inputs, hypothesis ...). Section 3 briefly outlines the dynamic programming theory and the algorithms used to solve this problem (quantization, backward dynamic programming,...). Section 4 presents some promising results obtained on simple cases with one and two targets.

TABLE OF CONTENTS

1	INTRODUCTION	1
2	PROBLEM STATEMENT	1
3	STOCHASTIC OPTIMAL CONTROL	2
4	RESULTS	4
5	CONCLUSION	5
	REFERENCES	6
	BIOGRAPHY	6

1. INTRODUCTION

Optimal control is a wide and interesting theory that appears in many fields such as, for instance, engineering, computer science, economics, operation research ... This paper aims to explain how to compute optimal trajectories for underwater vehicles evolving in a given environment to accomplish some tasks. This is an optimal control problem. Nevertheless, in real context, available inputs are not perfectly known. Hence a stochastic approach seems to be needed.

Markov decision processes (MDPs) constitute a general family of controlled stochastic processes suitable for the modeling of sequential decision-making problems. A significant list of references on discrete-time MDPs may be found in the survey [1] and the books [2], [3], [4], [5], [6], [7], [8]. The analysis of MDPs leads to mathematical and computational problems. The corresponding theory has reached a rather high degree of maturity, although the classical tools (such as value iteration, policy iteration, linear programming, and their various extensions) are generally hardly applicable in practice. This is mainly because MDPs are generally very large due to their inherent structure and, for instance, solving the associated dynamic programming equation leads to the

2. PROBLEM STATEMENT

Let us consider a general situation with a submarine of interest S (blue in Fig.1) surrounded by several targets (other submarines or surface ships). This submarine carries one or several sensors such as Cylindrical Array (CA), Flank Array (FA) or towed Array (TA).

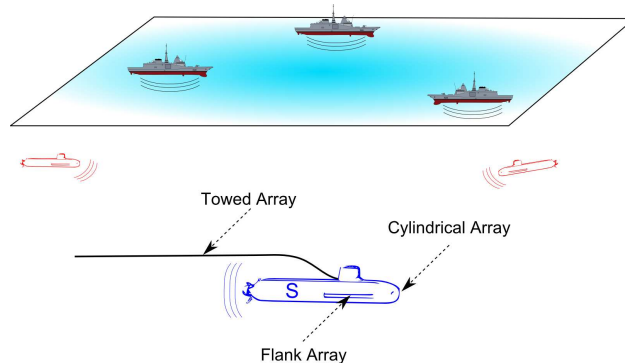


Figure 1. General situation.

The submarine S trajectory has to be controlled in order to satisfy the mission's objectives. These can be optimizing the different targets' detection range, minimizing its own detection range perceived by the other targets, reaching a way-point with minimum fuel consumption, and so on.

This objective is represented in a mathematical standpoint by the objective cost function J which is to be minimized by the control algorithm. It is supposed that we have the following inputs:

978-1-4577-0557-1/12/\$26.00 ©2012 IEEE.

¹Proc. 2012 IEEE/AIAA Aerospace Conf. Big Sky, MT, March 2012

- (1) S position and velocity,
- (2) Targets' position, course and speed, estimated by the Target Motion Analysis (TMA) function, see for instance [9].
- (3) Information about the environment (sound speed, sea floor depth,...),
- (4) A sound propagation code.

In complex situations, the submarine must fulfill several conflicting missions against multiple targets. For example, the submarine may have the objective of simultaneously increase the detection range of one target while keeping its own detection range as low as possible versus another target. In this case, the objective function J is a multi criteria aggregation function whose minimization will be a trade-off between these conflicting objectives.

Underwater sound propagation must be estimated on line. Sound propagation in the sea depends on a lot of environment parameters such as temperature, local pressure and salinity. Indeed, small variation of these parameters greatly modify sound propagation. The two figures Fig.2 and Fig.3 give examples of a sound velocity profile (sound speed versus depth) and a sound propagation diagram of an emitter source at a 300m depth (the third dimension, the color, being the signal level in dB).

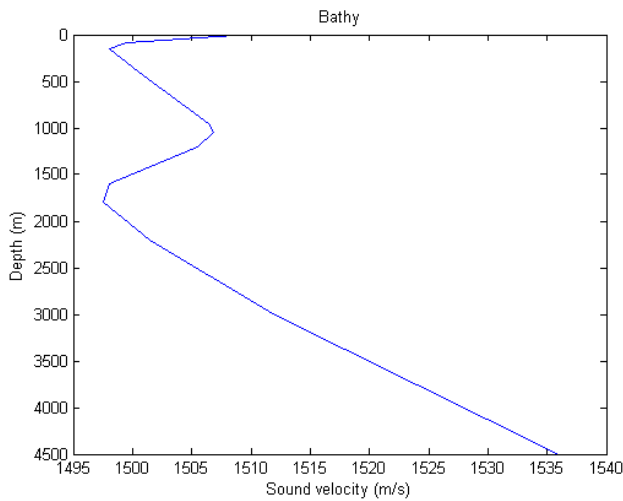


Figure 2. Sound velocity profile (depth vs velocity).

According to the objectives of S and previously described given inputs, the aim is to compute an optimal trajectory by applying a command $u(t)$ to S 's future states (see Fig.4).

This is an optimal control problem which aims to find a command $u(t)$ in order to minimize the cost function J .

3. STOCHASTIC OPTIMAL CONTROL

As seen in part 2, the problem is quite complex because there are uncertain inputs and we use a sound propagation code as a black box (with possible numerical problems, ...). Hence no strong assumptions (derivability, ...) can be made about this latter one. Due to these particularities, a discrete stochastic optimal control framework seems to be the best candidate to solve this problem. More precisely, we use a discrete-time finite horizon dynamic programming approach.

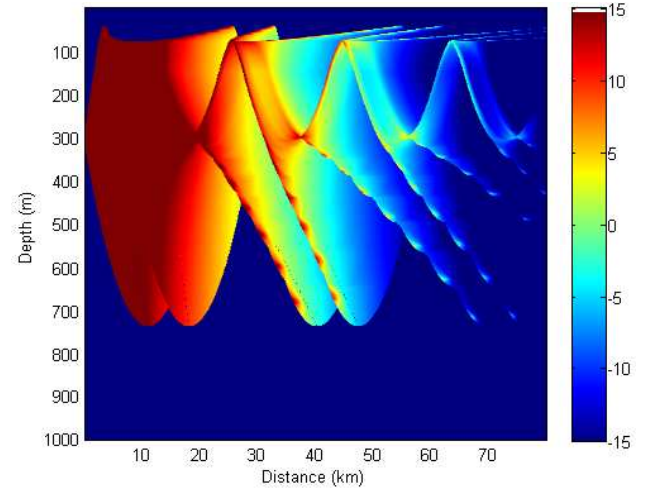


Figure 3. Sound propagation (depth distance signal level).

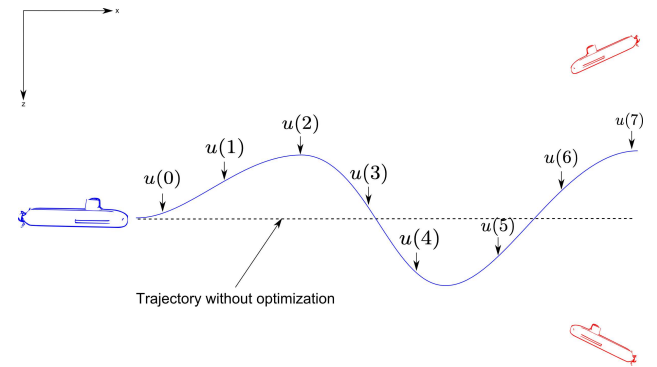


Figure 4. Optimal trajectory example.

Brief review of dynamic programming

In this section, we briefly introduce the discrete-time finite horizon Markov control model we are concerned with. Let us consider the following model :

$$(\mathcal{X}, A, \{A(x)|x \in \mathcal{X}\}, Q, c, C_N) \quad (1)$$

with

- \mathcal{X} : a Borel space, namely the state space
- A : a Borel space representing the control or action set
- $\{A(x)|x \in \mathcal{X}\}$: a family of non empty subsets of A , where $A(x)$ is the set of feasible controls or actions when the system is in state $x \in \mathcal{X}$. We suppose that $K = \{(x, a) \in \mathcal{X} \times A | a \in A(x)\}$ is a measurable subset of $\mathcal{X} \times A$.
- Q : a stochastic kernel on \mathcal{X} given K which stands for the transition probability function.
- $c : K \rightarrow \mathbb{R}$: is a measurable function representing the cost per stage.
- $C_N : \mathcal{X} \rightarrow \mathbb{R}$: is a measurable function representing the terminal cost.

Next, we define our Markov control model. Suppose that a finite horizon $N \geq 1$ and an initial state $x \in \mathcal{X}$

are given. The total expected cost of a policy $\pi = \{a_t | t = 0, 1, \dots, N-1\} \in \Pi$ (Π represents all possible policies) is defined as:

$$J(x, \pi) = E_x^\pi \left[\sum_{t=0}^{N-1} c(x_t, a_t) + C_N(x_N) \right] \quad (2)$$

$$\triangleq E \left[\sum_{t=0}^{N-1} c(x_t, a_t) + C_N(x_N) | \pi, x_0 = x \right] \quad (3)$$

The optimal total expected cost function is then defined as :

$$J^* = \inf_{\pi \in \Pi} J(x, \pi), \forall x \in \mathcal{X} \quad (4)$$

And we say that $\pi^* \in \Pi$ is an optimal policy if $J(x, \pi^*) = J^*(x)$ for every initial state $x \in \mathcal{X}$.

Hypothesis for our problem

In our application, we consider that the state space \mathcal{X} is finite. Hence the stochastic kernel Q is

$$P_{xy}^t(a) = \mathbb{P}(x_{t+1} = y | x_t = x, a) \quad (5)$$

Which is the probability that the system is in state y at time $t+1$ given that it was in state x at time t for a given control a .

Now consider $\{C_N, \forall x \in \mathcal{X}\}$ the final cost function and $c(x, a, y)$ the cost per stage function from state x to state y with a command a . The cost function (3) is then

$$J(\pi, x) = E \left[\sum_{t=0}^{N-1} c(x_t, a_t, x_{t+1}) + C_N(x_N) | \pi, x_0 = x \right] \quad (6)$$

The goal is to find an optimal policy that minimizes this cost. The main idea of the dynamic programming principle is the following algorithm

- $J_N^*(x) = C_N(x), \forall x \in \mathcal{X}$
- for $t \in \{N-1, \dots, 0\}$ and $\forall x \in \mathcal{X}$

$$J_t^* = \min_{a \in A(x)} \sum_{y \in \mathcal{X}} P_{xy}^t(a) [c(x, a, y) + J_{t+1}^*(y)] \quad (7)$$

$J_t^*(x)$ is the solution. Indeed, the optimal policy $\pi^* = \{a_0^*, a_1^*, \dots, a_N^*\}$ with

$$a_t^* = \arg \min_{a \in A(x)} \sum_{y \in \mathcal{X}} P_{xy}^t(a) [c(x, a, y) + J_{t+1}^*(y)] \quad (8)$$

is recursively constructed.

Submarine S

Suppose we have a submarine that can only control its depth. The state model is then:

$$z_{t+1} = z_t + a_t \quad (9)$$

At each time step, available controls are

$$a_t \in A(z) = \{-L\Delta_z, (-L+1)\Delta_z, \dots, -\Delta_z, 0, \dots, L\Delta_z\} \quad (10)$$

Where L is the possible number of depth steps from a state to another state and Δ_z the size of this step. z_t has discrete values and evolves in a growing finite grid Z_t

$$\forall t, z_t \in \{z_0 + n\Delta_z\} = Z_t, \quad (11)$$

$$n \in \{-tL, -tL+1, \dots, -1, 0, 1, \dots, tL\} \quad (12)$$

Target

We also suppose that the target evolves with a constant depth and its behavior is independent of S. The target is represented by a state vector composed by the submarine to target distance and the target speed $X_t = (d_t, s_t)$. The target position and speed are given by the TMA algorithms and the sate evolution is described by the stochastic model

$$X_{t+1} = FX_t + v_t \quad (13)$$

where $X_0 \rightsquigarrow \mathcal{N}(\mu_0, \Sigma_0)$, $F = \begin{pmatrix} 1 & \Delta_t \\ 0 & 1 \end{pmatrix}$, $v_t = (\epsilon_t^1, \epsilon_t^2) \rightsquigarrow \mathcal{N}(0, \Sigma_\epsilon)$. μ_0, Σ_0 and Σ_ϵ are known.

Solving a stochastic optimal control problem in a real context is challenging. In order to apply dynamic programming described in part 1, it is necessary to approximate \mathcal{X} with a finite space by using relevant methods.

Various methods already exist for this approximation (Kushner [10], Quantization [11], ...). We use a quantization method in the sequel because this method is powerful and also because convergence results are available.

Quantization

In our case, the process $X_t \in \mathbb{R}^2$. We approximate this process by a finite Markov chain thanks to the quantization method [11] :

- \hat{X}_t is the quantization of X_t and $\hat{\Gamma}_t$ is the M points grid at time t which represent X_t ($\hat{X}_t = (\hat{d}_t, \hat{s}_t) \in \hat{\Gamma}_t$)
- P_t^{ij} is the transition matrix, $P_t^{ij} \approx \mathbb{P}(\hat{X}_{t+1} = j | \hat{X}_t = i)$, $\forall (i, j) \in \hat{\Gamma}_t \times \hat{\Gamma}_{t+1}$.

The system at time t is $x_t := (z_t, \hat{X}_t) \in \mathbb{R}^3$. Component z_t is deterministic and \hat{X}_t is stochastic. Thanks to this quantization, we have a MDP with finite number of states. We have hence constructed the finite \mathcal{X} space described at the beginning of this section.

As the target and the submarine are supposed independent, the quantization of X_t can be done "off-line". The dynamic programming equation then becomes

$$J_t^*(x) = \min_{a \in A(x)} \sum_{y \in Z_t \times \hat{\Gamma}_t} P_t^{xy}(a) [c(x, a, y) + J_{t+1}^*(y)] \quad (14)$$

The process X_t is not controlled hence the transition matrix $P_t^{xy}(a)$ can be approximated by P_t^{ij} . $\forall x = (l, i) \in Z_t \times \hat{\Gamma}_t$ and $\forall y = (m, j) \in Z_{t+1} \times \hat{\Gamma}_{t+1}$

$$P_t^{xy}(a) \approx \mathbb{P}(z_{t+1} = l, \hat{X}_{t+1} = j) \quad (15)$$

$$\begin{aligned} z_t = m, \hat{X}_t = i, a_t = a & \quad (16) \\ = P_t^{ij} \mathbb{I}_{\{m=l+a\}} & \quad (17) \end{aligned}$$

That is why $\forall x = (l, i) \in Z_t \times \hat{\Gamma}_t$

$$\begin{aligned} J_t^*(l, i) &= \min_{a \in A(x)} \sum_{(m, l) \in Z_{t+1} \times \hat{\Gamma}_{t+1}} P_t^{ij} \mathbb{I}_{\{m=l+a\}} \\ &\quad [c((l, i), a, (m, j)) + J_{t+1}^*(m, j)] \\ &= \min_{a \in A(x)} \sum_{j \in \hat{\Gamma}_{t+1}} P_t^{ij} [c((l, i), a, (l+a, j)) \\ &\quad + J_{t+1}^*(m, j)] \end{aligned}$$

$c((l, i), a, (m, j))$ is the transition cost function from state (l, i) to state (m, j) given a control a . In our application, this function only depends on the sound propagation code which is “geometric” (values only depend on submarine and targets’s relative positions). Hence $c((l, i), a, (m, j)) = c(m, j)$

We finally obtain the following dynamic programming equation

$$J_t^*(l, i) = \min_{a \in A(x)} \sum_{j \in \hat{\Gamma}_{t+1}} P_t^{ij} [c(l+a, j) + J_{t+1}^*(l+a, j)] \quad (18)$$

$\forall x = (l, i) \in Z_t \times \hat{\Gamma}_t$. This final equation can be numerically solved because states are finite.

Numerical resolution

This problem is solved by a stochastic optimal control method. Our stochastic optimal control process is divided in three steps. The first step is an optimal quantization in order to approximate the target state by a finite Monte Carlo Markov chain state. The second step is a backward dynamic programming algorithm for evaluating the best policy from each possible system state. The last step is an in-line overall optimal control process which evaluates the best control sequence for our problem.

Step 1 : quantization

The goal is to approximate X_t by a finite Markov chain \hat{X}_t . We obtain for each time $t = 0, \dots, N$ a grid $\hat{\Gamma}_t$ and the corresponding transition matrices P_t^{ij} . Grids and transition matrices are stored.

Step 2 : dynamic programming

We compute in this part the optimal control that minimizes (18) for each element $(l, i) \in Z_t \times \hat{\Gamma}_t$ in the state space.

- Compute $\forall (l, i) \in Z_t \times \hat{\Gamma}_N$ thanks to $C_N(l, i)$ the final cost

$$\begin{cases} J_N^*(l, i) &= C_N(l, i) \\ a_N^*(l, i) &= 0 \end{cases} \quad (19)$$

- Compute for $t = N - 1, \dots, 0$ and $\forall (l, i) \in Z_t \times \hat{\Gamma}_t$

$$\begin{cases} J_t^*(l, i) &= \min_{a \in A(x)} \sum_{j \in \hat{\Gamma}_{t+1}} P_t^{ij} [c(l+a, j) \\ &\quad + J_{t+1}^*(l+a, j)] \\ a_t^*(l, i) &= \arg \min_{a \in A(x)} \sum_{j \in \hat{\Gamma}_{t+1}} P_t^{ij} [c(l+a, j) \\ &\quad + J_{t+1}^*(l+a, j)] \end{cases} \quad (20)$$

For each time and each state space point, following information are stored: $J_t^*(l, i)$ and $a_t^*(l, i)$. $a_t^*(l, i)$ is the best control starting from (l, i) at time t and $J_t^*(l, i)$ is the best cost at time t starting from (l, i) .

Step 3 : optimal control

This part is very simple. It is supposed that the trajectory of the target is known, say that the sequence of the target states is $\{X_0, \dots, X_N\}$. Then we can deduce the optimal control sequence from the optimal policy $\{a_t^*, t = 0, \dots, N\}$.

The single target case has been presented in detail in this section. The multitarget case is obtained in a similar way with a state vector containing two or more targets.

4. RESULTS

Single target case

We consider here a very simple scenario : the submarine is equipped with one sensor and there is a single target in the environment. The submarine wants to best detect the target regardless of its own detection range with respect to the other target. The initial geometry is depicted by the following Fig.5. The submarine S (blue) and the target

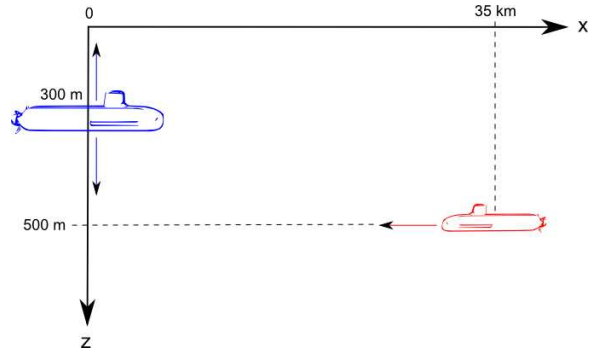


Figure 5. First scenario.

(red) have uniform motions hence the problem is to find an optimal depth for S (so, S motion is uniform only in the horizontal plane). The relative target’s speed is 10 m.s^{-1} . The environment (inputs of the sound propagation code) is described as follow:

- depth of the sea floor: 1000 m
- floor type : sand
- sea state: 3
- sound velocity profile: see (Fig.2)
- frequency: 1000 Hz

Results are illustrated² in Fig.6: The trajectory of the submarine S , in the depth-relative range plane, is plotted with a solid white line. The third dimension is the loss signal level seen from S . Results are in accordance with the

²Depth (m) versus distance (km)

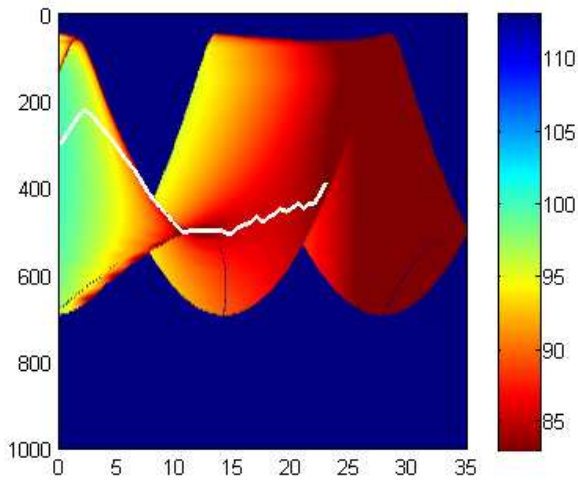


Figure 6. Results (initial depth: 300m).

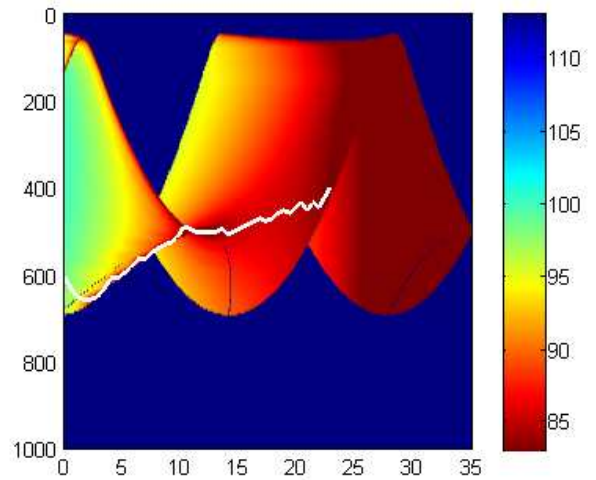


Figure 8. Results (initial depth: 600m).

good sense because the trajectory is always at depths where the target loss signal is minimum. Nevertheless, simple cases like this are interesting for a preliminary validation of complex algorithms.

Figures Fig.7 and Fig.8 are examples in the same conditions but with different initial depths for the submarine (100m and 600m).

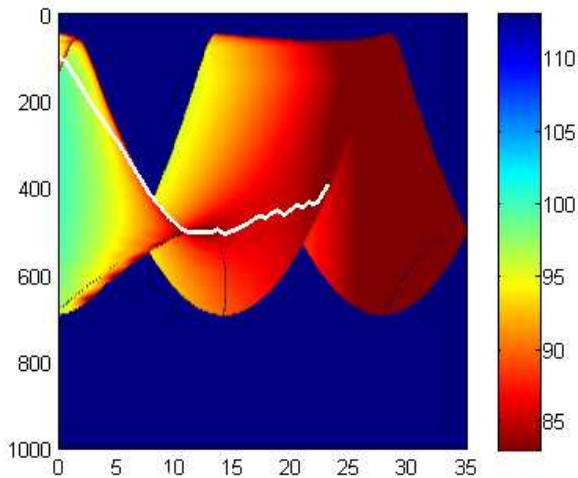


Figure 7. Results (initial depth: 100m).

The control seems to put the submarine at the right place where the loss of signal is minimum.

Multitarget case

Let us consider a more complex scenario with now two targets at different depth. The first target is still at 500m depth whereas the second target has an initial depth of 100m (see Fig.9).

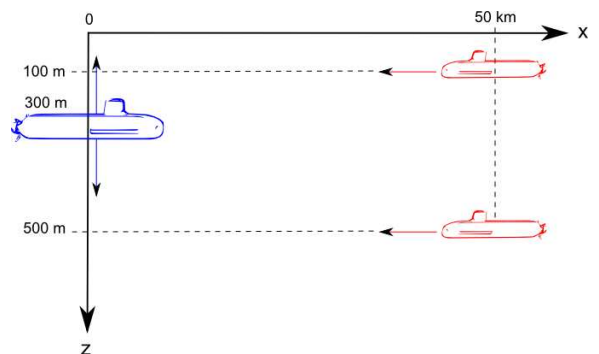


Figure 9. Second scenario.

The multitarget cost function is defined in order to keep a good detection range of each target. We obtain the following results illustrated on Fig.10 and Fig.11. Results are very promising because the trajectory of the submarine remains at a compromise depth where each target loss level is in a minimum area, though there is a little incursion in the blue with respect to target 1, but it corresponds to the edge of the best area with respect to target 2.

5. CONCLUSION

In this paper we have presented a stochastic optimal control framework which solves the difficult task of finding optimal trajectories for a submarine taking into account its mission objectives and the environment. First results obtained on simple scenarios show that the proposed framework is a very promising approach. Nevertheless, additional developments and further evaluations are needed.

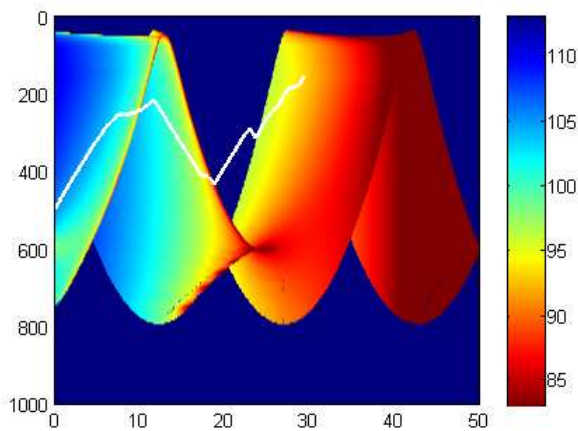


Figure 10. Results plotted on target 1 loss diagram.

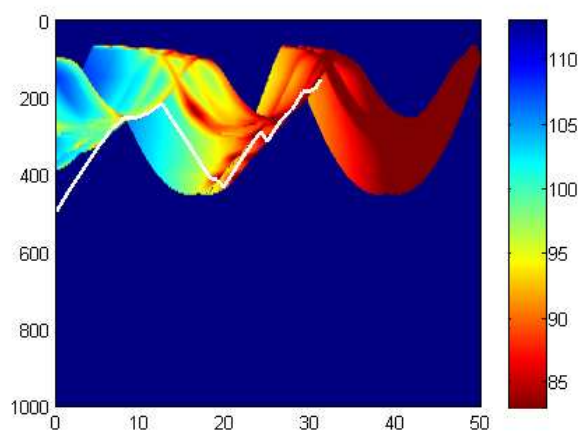


Figure 11. Results plotted on target 2 loss diagram.

REFERENCES

- [1] E. F.-G. M. G. A. Arapostathis, V.S. Borkar and S. Marcus, "Discrete-time controlled markov processes with average cost criterion : a survey," in *SIAM J. Control Optim.*, 31(2):282-344, 1993.
- [2] E. Altman, "Constrained markov decision processes," in *Stochastic Modeling. Chapman & Hall/CRC, Boca Raton*, 1999.
- [3] D. Bertsekas and S. Shreve, *Stochastic Optimal Control : the Discrete Time Case.* volume 139 of Mathematics in Science and Engineering. Academic Press Inc., New York, 1978.
- [4] J. Filar and K. Vrieze, *Competitive Markov Decision Processes.* Springer-Verlag, New York, 1997.
- [5] O. Hernandez-Lerma and J. Lasserre, *Discrete-time Markov Control Processes: Basic Optimality Criteria.* volume 30 of Applications of Mathematics, Springer-Verlag, New York, 1996.
- [6] —, *Further Topics on Discrete-time Markov Control Processes.* volume 42 of Applications of Mathematics, Springer-Verlag, New York, 1999.
- [7] M. Puterman, *Markov Decision Processes: Discrete Stochastic Dynamic Programming.* Wiley Series in

Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley & Sons Inc. New York, 1994.

- [8] L. Sennott, *Stochastic Dynamic Programming and the Control of Queuing Systems.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. Wiley & Sons Inc. New York, 1999.
- [9] D. Laneuville and C. Jauffret, *Recursive Bearings Only TMA via Unscented Kalman Filter: Cartesian vs Modified Polar Coordinates.* in Proc. 2008 IEEE/AIAA Aerospace conf. Big Sky, MT, March 2008.
- [10] H. Kushner and P. Dupuis, *Numerical Methods for Stochastic Control Problems in Continuous Time.* volume 24 of Applications of Mathematics, Springer-Verlag, New York, 1992.
- [11] G. Pages and H. Pham, *Optimal Quantization Methods for non Linear Filtering with Discrete-time Observations*, 2005.

BIOGRAPHY



Adrien Nègre received a bachelor degree in mathematics from Toulon university in 2005 and a "diplôme d'ingénieur" from the ENSIMAG (Ecole Nationale Supérieure d'Informatique et de Mathématiques Appliquées de Grenoble) in 2008. He joined DCNS group (2008) as a research engineer. His area of interest are Tracking, Image Processing and more recently Optimal Control



Olivier Marceau received the "diplôme d'ingénieur" from Ecole Supérieure d'Electricité (a.k.a. Supélec). He joined DCNS group (2008) as a research engineer. His area of interest are Tracking, Operations Research and more recently Optimal Control



Dann Laneuville received the "diplôme d'ingénieur" from Ecole Supérieure d'Electricité (a.k.a. Supélec) in 1987 and a Ph.D. in automatic and signal processing from the Paris XI Orsay University in 1998. He is currently with DCNS as a senior algorithms expert in Information and Surveillance Systems. Before joining DCNS in 2003, he has been with EADS for ten years and with MBDA for five years. His area of interest are Multitarget Tracking, Data Fusion, Target Motion Analysis and more recently Image Processing.



Huilong Zhang received the B.S. degree in applied mathematics from Northwestern Polytechnical University (China) in 1986 and a Ph.D in Applied Mathematics from University de Provence Aix Marseille 1 in 1992. He has been assistant professor at University Bordeaux I, he is also a member of the INRIA Team CQFD. His research interests include nonlinear filtering and

dynamic reliability.



Benoîte De Saporta received her Ph.D. in applied probability in 2004 from Rennes 1 University (France). Since 2006, she has been Assistant Professor at Montesquieu Bordeaux IV University, Department of Economics. She is also a member of the INRIA Team CQFD. Her research is focused on applied probability and more specifically estimation, control and numerical approximations

for autoregressive processes, Markov switching processes and hybrid systems, with applications in reliability and biology.



Francois Dufour is Professor at the Institut Polytechnique of Bordeaux (France) since 2002. He is also the leader of INRIA team CQFD (quality control and dynamic reliability). He is associate editor of SIAM Journal of Control and Optimization since. His research interests include Markov processes, stochastic stability and stochastic optimal control with applications in reliability and target tracking.

ability and target tracking.

- 4.2 [ZdSD⁺12] H. ZHANG, B. DE SAPORTA, F. DUFOUR, AND G. DELEUZE. Dynamic reliability : towards efficient simulation of the availability of a feedwater control system. In *NPIC-HMIT 2012*, San Diego, USA, 2012.

DYNAMIC RELIABILITY: TOWARDS EFFICIENT SIMULATION OF THE AVAILABILITY OF A FEEDWATER CONTROL SYSTEM

Huilong Zhang^{a,c,e}, Benoîte de Saporta^{b,d,e} François Dufour^{a,c,e}

^aUniv. Bordeaux, IMB UMR 5251, F-33400 Talence, France

^bUniv. Bordeaux, Gretha UMR 5113, F-33600 Pessac, France

^cCNRS, IMB, UMR 5251, F-33400 Talence, France

^dCNRS, Gretha, UMR 5113, F-33600 Pessac, France

^eINRIA Bordeaux Sud-Ouest, team CQFD, F-33400 Talence, France

huilong.zhang@math.u-bordeaux1.fr

Gilles Deleuze

EDF R&D, 1 Avenue du General de Gaulle, Clamart, France

ABSTRACT

In the reliability modeling of complex control systems, classical methodologies such as event-trees/fault-trees or Petri nets may not represent adequately the dynamic interactions existing between the physical processes (modeled by continuous variables) and the functional and dysfunctional behavior of its components (modeled by discrete variables). This paper proposes a framework for modeling and simulation of a water level control system in the Steam Generator (SG) of the secondary circuit of a nuclear power plant. We have developed a complete benchmark case. The behavioral model of SG is obtained from a linearized model published in 2000 by EDF [1,2]. Four physical variables (steam flow rate, water flow rate, steam-water level, water level) are modeled; they follow a system of linear differential equations with piecewise constant coefficients, coupled with a PID controller that regulates the water level in the SG. Detailed description of the components, failure modes and control laws of the principal components is presented. For modeling the system, we use the Piecewise Deterministic Markov Processes (PDMP) framework and for implementation we chose Simulink associated with Stateflow. PDMP offer a very general modeling framework to deal with dynamic reliability problems; Simulink is an appropriate tool to simulate non linear differential equations and their controller, while Stateflow implementation is appropriate for finite state machine descriptions of different components.

Key Words: Dynamic PRA/PSA, Piecewise Deterministic Markov Processes

1 INTRODUCTION

Hybrids systems are described by continuous variables, deterministic events and stochastic events (e.g. control logic and mechanical parts failures, unplanned variations of operational profile...). For a large class of industrial processes, the layout of operational or accidental sequences generally comes from the occurrence of two types of events:

- the first type is directly linked to a deterministic evolution of the physical parameters of the process,
- the second type of events is purely stochastic. It usually corresponds to random demands or failures of system components.

Unfortunately, the static methods used for systems reliability modeling, such as combinatory approaches (fault trees, event trees, reliability diagrams) are not relevant to model hybrid systems. This is a current challenge in reliability

analysis, and requires the development of so-called “Integrated Deterministic-Probabilistic Safety Analysis (IDPSA). The need of IDPSA methods comes from the observation that static methods applied in PSA are limited to find time dependent interactions and unknown vulnerable sequences regarding physical phenomena, control logic, operator actions, equipment failures [8].

The benchmark system that we study is a Feedwater Control System (FCS) for the Steam Generator (SG) in the secondary circuit of a nuclear plant. The mission of the system is to maintain the water level in the steam generator around a reference position. The mission fails if the water level rises above or falls beyond threshold limits. This test case has the advantage of being representative of a real system and to cover most of the situations encountered in the dynamic reliability literature. A similar benchmark system was described by the U.S. Nuclear Regulatory Commission [4] where two approaches for dynamic reliability were compared: DFM (Dynamic Flowgraph Methodology) and Markov/CCMT (Cell-to-Cell Mapping Technique). But the date released by the NRC report did not permit us to reconstruct the model.

The work presented here is a continuation of a series of works already realized within the INRIA team CQFD. They are intended to illustrate the efficiency of a method combining the modeling power of Piecewise Deterministic Markov processes (PDMP) and the Monte Carlo computational implementation, to treat certain dynamic reliability problems. We have in the past modeled and simulated systems of "academic" size [6] and "industrial" size [5], where implementations were done in C++ or Matlab. But to model the Feedwater Control System, we chose the software Simulink/Stateflow of Mathworks. This paper is organized as follows. Section 2 briefly presents the Piecewise Deterministic Markov processes (PDMP). Section 3 gives the functional description of our benchmark. Section 4 then presents the implementation by Simulink/Stateflow. Section 5 shows some numerical results. Finally, section 6 presents our conclusions.

2 PIECEWISE DETERMINISTIC MARKOV PROCESSES

PDMP provide a very general modeling framework to deal with dynamic reliability problems. Let M be the finite set of the possible regimes of the system. For all m in M , let E_m be an open set of \mathbb{R}^d . A Piecewise deterministic Markov process is defined from the three local characteristics (Φ, λ, Q) where

- the flow $\Phi : M \times \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}^d$ is continuous and for all $s, t \geq 0$, $\Phi(\bullet, t+s) = \Phi(\Phi(\bullet, s), t)$. It describes the trajectory of the deterministic process between jumps. For all (m, x) in $M \times E_m$, we set

$$t^*(m, x) = \inf \{ t > 0 : \Phi(m, x, t) \in \partial E_m \}$$

the time to reach the boundary of the domain.

- the jump intensity λ characterizes the frequency of jumps. For all (m, x) in $M \times E_m$, and $t \leq t^*(m, x)$, we set

$$\Lambda(m, x, t) = \int_0^t \lambda(\phi(m, x, s)) ds$$

- the Markov kernel Q represents the transition measure of the processes that allows to select the new position after each jump.

The trajectory $X_t = (m_t, x_t)$ of the process can be defined iteratively. We start from an initial point $X_0 = (k_0, y_0)$ with $k_0 \in M$ and $y_0 \in E_{k_0}$. The first jump time T_1 is determined by the distribution

$$P_{(k_0, y_0)}(T_1 > t) = \begin{cases} e^{-\Lambda(k_0, y_0, t)}, & t < t^*(k_0, y_0) \\ 0, & t \geq t^*(k_0, y_0) \end{cases}$$

On the interval $[0, T_1)$, the process follows the deterministic trajectory $m_t = k_0$ and $x_t = \Phi(k_0, y_0, t)$. At the random time T_1 , the process has a jump. The regime changes and the process is then reset at X_{T_1} , a random variable that follows the law given by $Q_{k_0}(\Phi(k_0, y_0, T_1), \bullet)$. We then similarly draw a new jump time $T_2 - T_1$, and on the interval $[T_1, T_2)$ the process follows the trajectory $m_t = k_1$ et $x_t = \Phi(k_1, y_1, t - T_1)$. This builds iteratively the PDP.

A particularity of the steam generator system is that it is regulated. The flow Φ is thus the solution of a differential equation controlled by a PID controller. It admits no analytical solution and will be numerically approximated at each time step. The reason why we choose the PDP model is twofold. First, it provides a modeling framework that is both general and accurate. Second, this model offers the perspective in the future to perform optimal control: optimal stopping, predictive maintenance [7], etc.

3 FUNCTIONAL DESCRIPTION

We have modeled a part of the secondary circuit of a pressurized water reactor. It is composed of seven components: one passive system representing the whole steam transport system (VVP), three extraction pumps (CEX), two feeding turbo pumps (TPA), and one water flow regulation valve (ARE). The rest of the secondary circuit does not interest us for the test case as it has no direct influence on the reliability and safety of the circuit. The reliability diagram we modeled is given in Figure 1.

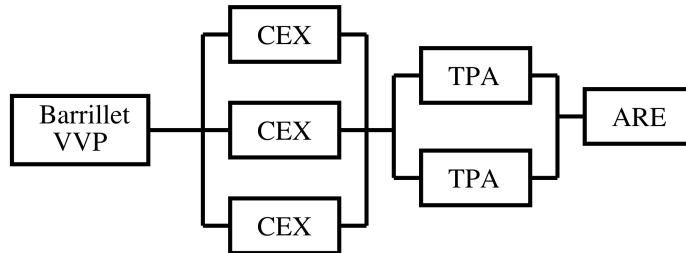


Figure 1. Reliability diagram for the mechanical and electromechanical systems of FCS.

3.1 Description of the system

The VVP barrel maintains the steam flow to the turbo pumps and dryers. A breakdown in the barrel VVP is a critical point of failure and represents a minimum level of system reliability. The three pumps CEX maintain the vacuum in the condenser (upstream of VVP) and ensure a flow of feeding water. They are redundant in 2/3. The third pump is stopped, in standby. It is started when one of the other pumps fails. A failed pump, once repaired, remains in standby. The two turbo pumps TPA work together, they provide the common pressure to the SG, discharged into a common cylinder integrated in the VVP. In the case of failure of a TPA, the second one switches over speed and provides some of the charge. We consider in this model that the power of the installation decreases automatically to 60% when only a single TPA works. Finally, the actuator (ARE) is used to command the feed water flow rate in the SG. It consists of a main valve and a bypass valve. A logic sequence determines the openings and closing of the individual valves as a function of the power of the installation.

In our benchmark system, four physical processes are considered: the feed water flow rate (Q_e), the steam flow rate (Q_v), the narrow range water level (N_{ge}) and the wide range water level (N_{gt}). A PID controller is used to maintain the water level within limits of reference-points. The behavioral model of the SG is obtained from a non-linear model published in 2000 by EDF [1, 2]. The general control strategy consists in maintaining the narrow range and wide range water levels within limits of their set points. This can be accomplished by concentrating the control effort on the single controlled variable: the narrow range water level Q_{ge} . Figure 2 shows the basic PID feedback control structure.

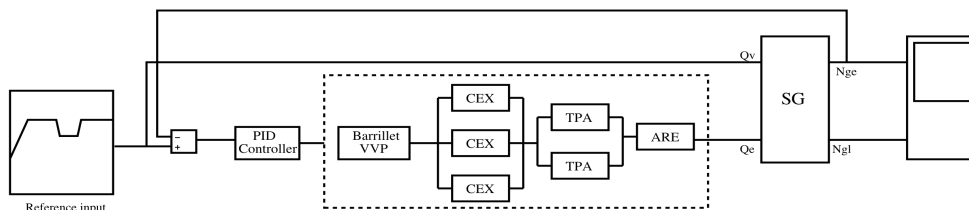


Figure 2. Basic feedback structure of the PID controller

3.2 Steam generator level control model

The model used in this study for the purpose of controller design, is a somewhat simplified version of the Irving model [1]. It was developed by the Research and Development division of EDF. The detailed model for the Steam Generator can be found in [2]. Such theoretical models use fundamental conservation equations for mass, energy, momentum, volume and basic thermodynamic principles. It is a simple fourth-order model. We will give a brief description. The relationship among N_{ge} , Q_e , Q_v and N_{gl} can be modeled by the following transfer functions:

$$N_{ge}(s) = \frac{1}{T_n s} \left(\frac{Q_e(s)}{(1 + \tau s)(1 + T_h s)} - \frac{1 - F_g T_g s}{1 + T_g s} Q_v(s) \right), \quad N_{gl}(s) = \frac{1}{T_{int} s} (Q_e(s) - Q_v(s)). \quad (1)$$

The parameters T_n, F_g, T_h, τ are functions of the operating power P_n and are summarized in Table I. Note that $T_g = 1.429$ and $T_{int} = 20$ are constants. The term $1 - F_g T_g s$ is incorporated to account for the modeling of the two-phase swell and shrink effects. Due to proprietary reasons, the model discussed in this paper is a scaled and modified version of the model used by EDF.

Table I. Parameters variation over the power range

P_n (%)	3.2	4.1	9.5	24.2	30	50	100
T_n	5.14	8.00	9.00	6.29	5.71	5.71	5.71
F_g	13.0	18.00	10.00	4.00	4.00	4.00	4.00
T_h	24.29	8.00	4.29	1.43	1.14	0.71	0.71
τ	1.43	1.43	1.43	4.29	4.29	4.29	4.29

Denoting the water levels by $y_1 = N_{ge}$ and $y_2 = N_{gl}$ and the steam and feed-water flow-rates by $d = Q_v$ and $u = Q_e$, we have following equivalent state-space form:

$$\dot{x}(t) = \begin{bmatrix} 0 & 0 & 0 & \frac{1}{T_n} \\ 0 & -\frac{1}{T_h} & 0 & -\frac{1}{T_n} \\ 0 & 0 & -\frac{1}{T_g} & 0 \\ 0 & 0 & 0 & -\frac{1}{\tau} \end{bmatrix} x(t) + \begin{bmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{\tau} \end{bmatrix} u(t) + \begin{bmatrix} -\frac{1}{T_n} \\ 0 \\ 1 + F_g \\ \frac{T_n}{T_n} \\ 0 \end{bmatrix} d(t),$$

$$y(t) = \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} \frac{1}{T_n} & 1 & 1 & 0 \\ \frac{1}{T_{int}} & 0 & 0 & \frac{\tau}{T_{int}} \end{bmatrix} x(t).$$

3.3 Reliability data and state graphs

Reliability data and state graphs have been defined for each component. We provide in Table II and Figure 3 the details for the simplest case, the VVP. The necessary data for the other sub systems are much more numerous and will not be presented in this paper due to obvious reasons of clarity. The worst case is that of the ARE with four graphs and more than a dozen states and thirty parameters.

Table II. Reliability data example (VVP) (Failure rate=2.17e-5/h)

Mesh	Contribution to the failure rate	Pfd	MTTR	Effect	Mode of failure
Outflow	89%	na	12	Automatic reactor trip	Mode I
Break	11%	na	168	Automatic reactor trip	Mode II

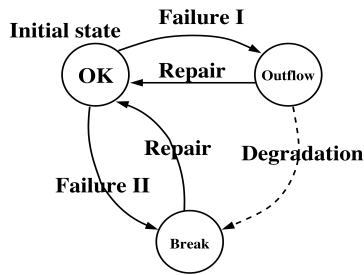


Figure 3. State graph of barrel VVP

3.4 Test scenario

We consider the following scenario, illustrated in Figure 4. After a piecewise linear rise lasting 24 hours, the system reaches its stationary state, 100% P_n and remains there for 18 months, followed by a descent of 24 hours. The objective is to simulate the behavior of the system subject to random failures. The simulation is stopped when an automatic reactor trip (RT) occurs. In this scenario, we suppose that the control law of the two ARE valves satisfies the following logic: when the operation power is in [0%, 2%], a supplement system (ASG) is applied to control the feed water rate, between [2% 15%], the bypass valve is used, and if P_n is above 15%, the bypass valve is turned off and the main valve comes into operation.

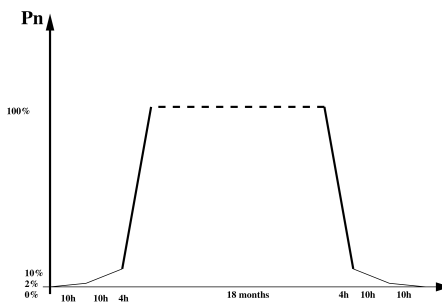


Figure 4. Test case scenario

4 SIMULINK/STATEFLOW IMPLEMENTATION

The global scheme of the simulator is presented in Figure 5. The *Ramp* block is parameterized to generate a constantly increasing or decreasing signal (P_n). The range of the ramp is restricted to [0,100] by the *saturation* block. Four output signals N_{ge} , P_n , N_{gl} and RT (automatic reactor trip) are computed, the last one stopping the simulation when an RT occurs.

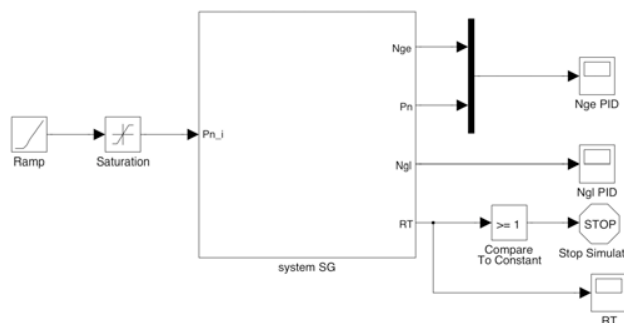


Figure 5. Global scheme of simulator

The Steam generator is fully modeled by the subsystem SG, presented in Figure 6. It consists of

- a Stateflow chart named *installation*, which includes the seven components VVP, CEX, TPA and ARE described above. This block is activated at each time step, and also whenever the power P_n crosses one of the thresholds 2%, 15%.
- the Steam Generator modeled by the *gv* block with two inputs (Q_v, Q_e) and two outputs (N_{ge}, N_{gl}). This system obeys a system of differential equations which coefficients depend on Q_v . Here the operation power P_n and steam rate Q_v share the same signal, because they are supposed to be proportional.
- a PID controller, which input is the difference between the set point P_n and N_{ge} , and which output is Q_e , the flow rate of feed water injected into the SG. The variable Q represents the disturbances from ARE.

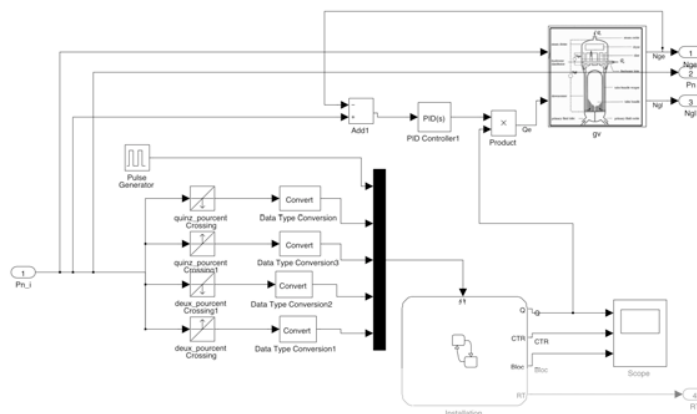


Figure 6. SG subsystem

A main advantage of Simulink/Stateflow modeling is that it takes the form of an interactive graph, which makes easy the understanding of the model. If the system works in nominal mode, when no component is down, the water level is controlled by the PID controller. If a component fails, it can either cause an AAR or a minor fault. In the former case the simulation is stopped, in the latter case the simulation goes on, the component is under repair, a decreasing followed by an increasing ramp is scheduled if necessary.

The Stateflow chart *installation* models the discrete behavior of the SG. It includes all the seven discrete components ARE, VVP, 3 CEX and 2 TPA, see Figure 7. We will present in detail here only two components: VVP and CEX.

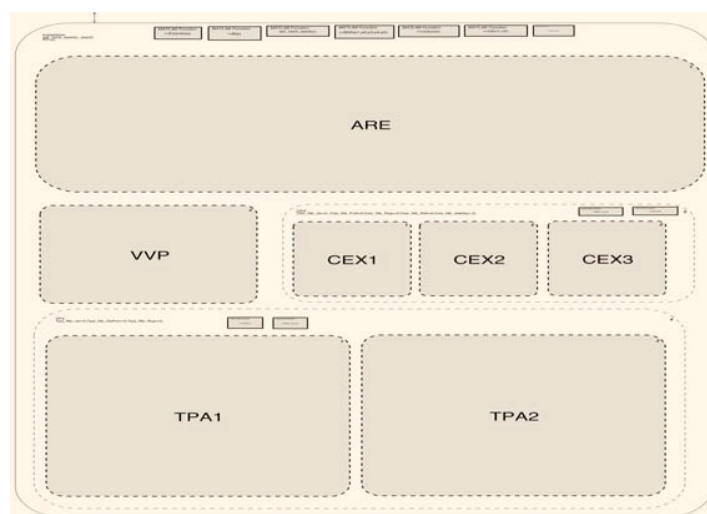


Figure 7. Stateflow chart for installation

4.1 VVP Modeling

This component has three possible states (*OK*, *Outflow*, *Break*), see Figure 8. The time spent in each state is exponentially distributed. When this component is active, its default state is *OK*. Two random variables are then drawn, $x=E(2.17e-5/h)$ represents an exponential distribution with parameter $\lambda=2.17e-5/h$, and $p=B(0.89)$ is a Bernoulli draw. When the time spent in this state exceeds x , the transition after(x , sec) occurs. The state of the component switches to *Outflow* if $p=1$ or to *Break* if $p=0$. In this latter case, an RT signal is sent to stop the simulation. The codes 301 and 302 can record the failed component and the cause of the failure, for future Monte Carlo analysis.

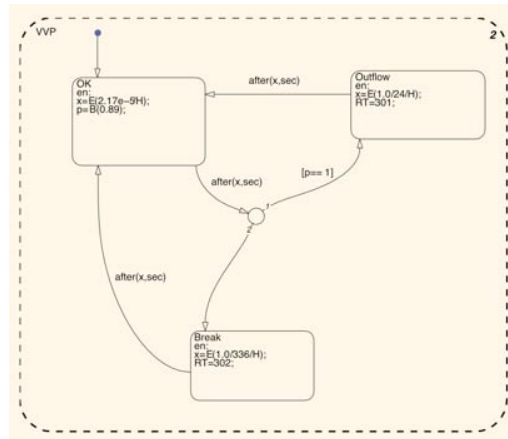


Figure 8. VVP Model

4.2 CEX Modeling

The principle of operation of the CEX is similar, but much more complex. Figure 9 illustrates the implementation of a CEX, this is indeed the details of CEX1 found in Figure 7. The other two CEX are almost identical (copy-paste), only the initial conditions are different. In the initial state, two CEX are in operation, the third is in standby.

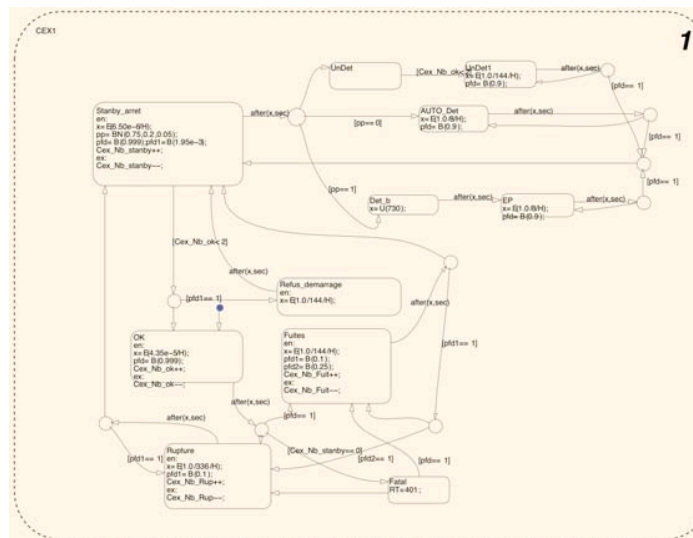


Figure 9. CEX implementation

5 NUMERICAL RESULTS

The total duration of the scenario is 18 months. There are two types of regimes: transient and steady. When the system is in the transient regime, the water level N_{ge} and power P_n vary rapidly. To follow this command, one must choose a discretization time step small enough (0.6 seconds), otherwise the PID controller loses its stability. In the steady state, all physical variables remain constant, only the component failures affect the system state. As the system is very reliable, the nominal duration is often very long and a small time step is not pertinent because it dramatically slows down the simulator. We propose a specific technique to solve this problem by using inhomogeneous time steps. Two distinct Simulink models were created, the first one with the PID, and the second one without PID. This allows setting two different time steps. During the transient period, we use a time step of 0.6 seconds, and during the stationary period, a time step of 60 minutes is sufficient to simulate component failure.

To illustrate the results, we simulated a history without failure, and we set steady period to be 3 days (instead of 18 months, for illustrative reason). Figure 10 shows that the N_{ge} (red) coincides well with the reference point (blue).

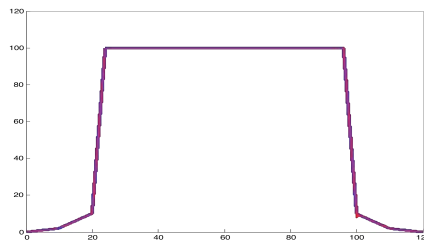


Figure 10. Reference point and N_{ge} level

On a total of 4000 simulated histories, 2190 Reactor Trips (RT) occurred, representing a probability of 54.75%. That is to say that in this scenario the system has about one chance over two to suffer a RT, for a period of 18 months. Table III summarizes the number of RT caused by each component. It also gives the percentage of occurrence among the 2190 RT.

Table III. Number of RT over 4000 histories

Sub system	Number of RT	Percentage
VVP	792	36%
ARE	1301	59%
CEX	50	2.28%
TPA	47	2.1%
Total	2190	100%

We found no RT caused by the bypass valve ARE. This can be explained by the fact that the stay in the transitional regime is too short (48 hours) compared to the total duration (18 months) of the scenario. This failure is a rare event. We also note that many RT are caused by the barrel VVP (36%). Its failure rate ($2.17e-5/h$) is comparable to those of the CEX ($4.35e-5$) and TPA ($5.9e-4$), but these components are redundant unlike the VVP, which minimizes the RT.

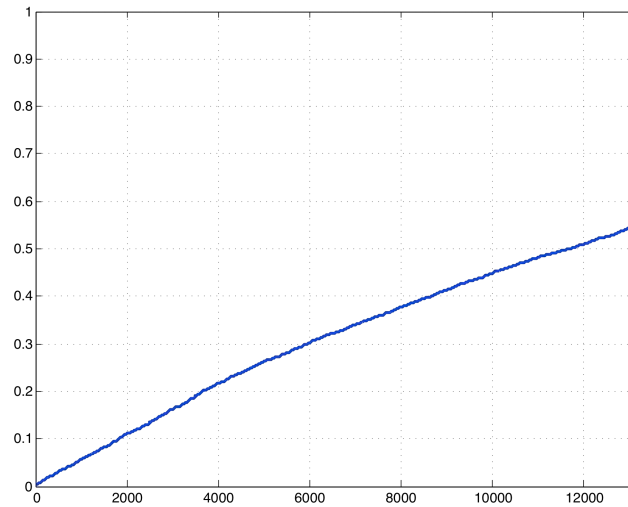


Figure 11. Cumulative probability of RT

Figure 11 illustrates the cumulative probability of Reactor Trips over time.

6 CONCLUSION

The modeling by PDMP applies very well to this problem of dynamic reliability. The approach combined with Simulink/Stateflow allows building an interactive simulator. It therefore offers interesting perspectives in several points of view.

- Graphical programming. The source code looks like a reliability diagram. In debugger mode, users can view the states and transitions step by step.
- Upgrade maintenance of the simulator. Components VVP, CEX, TPA, ARE can be modeled and tested separately (assuming that the other components are 100% reliable) and then combined by simple copy and paste. We can easily add in the future other components. Similarly, one can handle the problem of redundant components in pre-building a component library.
- Limiting the number of components. One of the main difficulties in hybrid system modelling, is the number of possible modes. By using this approach, there is no problem of combinatorial explosion. Indeed, the state machine of Stateflow is component oriented, that is to say that, at each time step and for each component, the simulator calculates the component state separately. It is not necessary to know a priori the number of possible states.

The main disadvantage of this approach is execution time. For the test case that we handled, a story is simulated in about 30 seconds (on a laptop), so 33 hours are necessary to run 4000 Monte Carlo runs. Experience shows that a C++ simulator dedicated to a problem of this size can probably run ten times or hundred times faster, but at the cost of a heavy investment in programming and the generated code is hardly evolutionary [5,6]. We have partially solved the problem by using the parallel computing toolbox of Mathworks. A computer equipped with 12-cores, reduced the computation time to 3 hours.

Others perspectives can also be explored. Representing the positive feedback between system trips and reliability or of mechanical systems or sensors drift will be a step towards dynamic reliability representation. Also, from the basis of this model, simulations may be used to simulate fault inside the control logic, incl. measurement logic. Moreover, the INRIA CQFD team proposes numerical algorithms for optimal control: optimal stopping, impulse control, etc [7], that permit to study various inspection and maintenance strategies for the FCS, taking into account hybrid and

dynamic aspects. In all these methods, the Monte Carlo simulator is the key step necessary to perform the optimization.

7 ACKNOWLEDGMENTS

This work was financed by the GIS 3SGS (Groupement d'Interêt Scientifique, Surveillance, Sûreté et Sécurité des Grands Systèmes) under contract APPRODYN (APPROches de la fiabilité DYNamique pour modéliser des systèmes critiques).

8 REFERENCES

1. E. Irving, C. Miossec and J. Tassart, "Towards efficient full automatic operation of the PWR steam generator with water level adaptive control", *Proceedings of the International Conference on Boiler Dynamics and Control in Nuclear Power Stations*, British Nuclear energy Society, London, pp. 309-329, (1980).
2. M.V. Kothare, B. Mettler, M. Morari, P. Bendotti and C.M. Falinower, "Level control in the steam generator of a nuclear power plant". *IEEE Transactions on Control Systems Technology*, Vol. 8, pp. 55-69 (2000)
3. M.H.A. Davis, *Markov models and optimization*, Chapman and Hall, London, (1993).
4. T. Aldemir, S. Guarro, et al., "Benchmark Implementation of Two Dynamic Methodologies for the Reliability Modeling of Digital Instrumentation and Control Systems", NUREG/CR-6985, US Nuclear Regulatory Commission, Washington, (2008).
5. H. Zhang, F. Dufour, Y. Dutuit and C. Elegbede, "Application des processus déterministes par morceaux à un système de production pétrolière offshore", *Proceedings of $\lambda\mu$ -16*, Avignon, France, (2008).
6. H. Zhang, F. Dufour, Y. Dutuit and K. Gonzalez, "Piecewise deterministic Markov processes and dynamic reliability", *Journal of Risk and Reliability*, Vol 4, pp. 545-551, (2009)
7. B. De Saporta, F. Dufour and K. Gonzalez, "Numerical method for optimal stopping of piecewise deterministic Markov processes", *Annals of Applied Probability*, n° Vol 5, pp.1607-1637, (2010).
8. Y. Adolfsson, J.E. Holmberg, G. Hultqvist, P. Kudinov, I. Männistö, "Proceedings of the Deterministic/probabilistic safety analysis workshop October 2011", *Research report VTT-R-07266-11*, 2011.