From algebraic flux correction schemes to nonlinear edge diffusion

Gabriel R. Barrenechea[1]

[1] Department of Mathematics and Statistics, University of Strathclyde, Scotland

Advanced numerical methods:
Recent developments, analysis and applications
Institut Henri Poincaré, Paris
October 3-7, 2016

University of Strathclyde Glasgow

This talk gathers contributions made in collaboration with:

1. Alejandro Allendes (UTFSM, Chile)
2. Erik Burman (UCL, UK)
3. Volker John (WIAS, Germany)
4. Fotini Karakatsani (Chester, UK)
5. Petr Knobloch (Charles University, Czech Republic)
6. Richard Rankin (UTFSM, Chile)

University of Strathclyde Glasgow

Outline

1. Motivation: From AFC schemes to nonlinear edge diffusion.
2. The edge diffusion method: Main properties.
3. An a posteriori error estimator.
4. Numerical results.
5. Concluding remarks.

University of Strathclyde Glasgow

Introduction: The discrete maximum principle

The continuous maximum principle :

Theorem

Let u be the solution of the problem

$$-\varepsilon\Delta u + \boldsymbol{b}\cdot\nabla u + \sigma u = f \quad \text{in}\,\Omega\,,$$

and $u = 0$ on $\partial\Omega$. Then, if $f \geq 0$ in Ω, u can not attain a negative interior minimum in Ω.

University of Strathclyde Glasgow

The discrete maximum principle

The DMP :

Theorem

Let $u_h \in \mathbb{P}_1(\Omega)$ be the solution of the problem

$$\varepsilon\,(\nabla u_h, \nabla v_h)_\Omega + (\boldsymbol{b} \cdot \nabla u_h, v_h)_\Omega + (\sigma u_h, v_h)_\Omega = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

Then, if $f \geq 0$ in Ω, the mesh is acute, and $\frac{|\boldsymbol{b}|h + \sigma h^2}{2\varepsilon} < 1$, then u_h does not attain a negative minimum in Ω.

Remark : Under these hypothesis, the matrix

$$[\varepsilon(\nabla \psi_j, \nabla \psi_i)_\Omega + (\boldsymbol{b} \cdot \nabla \psi_j, \psi_i)_\Omega + (\sigma \psi_j, \psi_i)_\Omega]$$

is an M-matrix. This is, it is invertible, all the diagonal elements are positive, and the off-diagonal ones are non-positive.

The discrete maximum principle

The DMP :

Theorem

Let $u_h \in \mathbb{P}_1(\Omega)$ be the solution of the problem

$$\varepsilon\,(\nabla u_h, \nabla v_h)_\Omega + (\boldsymbol{b}\cdot\nabla u_h, v_h)_\Omega + (\sigma u_h, v_h)_\Omega = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

Then, if $f \geq 0$ in Ω, the mesh is acute, and $\frac{|\boldsymbol{b}|h+\sigma h^2}{2\varepsilon} < 1$, then u_h does not attain a negative minimum in Ω.

Remark : Under these hypothesis, the matrix

$$[\varepsilon(\nabla\psi_j, \nabla\psi_i)_\Omega + (\boldsymbol{b}\cdot\nabla\psi_j, \psi_i)_\Omega + (\sigma\psi_j, \psi_i)_\Omega]$$

is an M-matrix. This is, it is invertible, all the diagonal elements are positive, and the off-diagonal ones are non-positive.

Algebraic flux correction schemes

Starting point : A finite element discretisation of our problem of the form:

$$\mathbb{A}\mathrm{U} = \mathrm{G}\,.$$

Define:

$$\mathbb{D} := (d_{ij}) \quad \text{where} \quad d_{ij} := -\max\{a_{ij}, 0, a_{ji}\} \text{ for } i \neq j\,, \quad d_{ii} = -\sum_{j \neq i} d_{ij}\,.$$

Remark: The matrix $\tilde{\mathbb{A}}$ is an M-matrix. Then, it preserves positivity.

Algebraic flux correction schemes

Starting point : A finite element discretisation of our problem of the form:

$$\mathbb{A}\mathrm{U} = \mathrm{G}\,.$$

Define:

$$\mathbb{D} := (d_{ij}) \quad \text{where} \quad d_{ij} := -\max\{a_{ij}, 0, a_{ji}\} \text{ for } i \neq j\,, \quad d_{ii} = -\sum_{j\neq i} d_{ij}\,.$$

Remark: The matrix $\tilde{\mathbb{A}}$ is an M-matrix. Then, it preserves positivity.

Algebraic flux correction schemes

Starting point : A finite element discretisation of our problem of the form:

$$(\mathbb{A}+\mathbb{D})\mathrm{U} = \mathrm{G} + \mathbb{D}\mathrm{U}\,.$$

Define:

$$\mathbb{D} := (d_{ij}) \quad \text{where} \quad d_{ij} := -\max\{a_{ij},0,a_{ji}\} \text{ for } i\neq j\,, \quad d_{ii} = -\sum_{j\neq i} d_{ij}\,.$$

Remark: The matrix $\tilde{\mathbb{A}}$ is an M-matrix. Then, it preserves positivity.

Algebraic flux correction schemes

Starting point : A finite element discretisation of our problem of the form:

$$\big(\underbrace{\mathbb{A}+\mathbb{D}}_{=:\tilde{\mathbb{A}}}\big)\mathrm{U} = \mathrm{G} + \mathbb{D}\mathrm{U}\,.$$

Define:

$$\mathbb{D} := (d_{ij}) \quad \text{where} \quad d_{ij} := -\max\{a_{ij}, 0, a_{ji}\} \text{ for } i \neq j\,, \quad d_{ii} = -\sum_{j\neq i} d_{ij}\,.$$

Remark: The matrix $\tilde{\mathbb{A}}$ is an M-matrix. Then, it preserves positivity.

Algebraic flux correction schemes

Equivalent system :

$$\tilde{\mathbb{A}}\,\mathrm{U} = \mathrm{G} + \mathbb{D}\mathrm{U}\,.$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.

Algebraic flux correction schemes

Equivalent system :

$$\tilde{\mathbb{A}}\,\mathrm{U} = \mathrm{G} + \mathbb{D}\mathrm{U}\,.$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.

University of Strathclyde Glasgow

Algebraic flux correction schemes

Equivalent system :

$$(\tilde{\mathbb{A}}\,\mathrm{U})_i = g_i + \sum_{j\neq i} f_{ij}$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.
The limiters α_{ij} should satisfy the following:

- $\alpha_{ij} \in [0,1]$;
- α_{ij} should be as close to 1 as possible;
- $\alpha_{ij} \approx 1$ where the Galerkin solution is smooth.
- $\alpha_{ij} = \alpha_{ji}$.

Algebraic flux correction schemes

Equivalent system :

$$(\tilde{\mathbb{A}}\,\mathrm{U})_i = g_i + \sum_{j\neq i} f_{ij}$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.

The limiters α_{ij} should satisfy the following:

- $\alpha_{ij} \in [0,1]$;
- α_{ij} should be as close to 1 as possible;
- $\alpha_{ij} \approx 1$ where the Galerkin solution is smooth.
- $\alpha_{ij} = \alpha_{ji}$.

Algebraic flux correction schemes

Equivalent system :

$$(\tilde{\mathbb{A}}\,\mathrm{U})_i = g_i + \sum_{j\neq i} \alpha_{ij}(U) f_{ij}$$

From the properties of $\mathbb{D}$ it follows that

$$\left(\mathbb{D}\mathrm{U}\right)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.

The limiters α_{ij} should satisfy the following:

- $\alpha_{ij} \in [0,1]$;
- α_{ij} should be as close to 1 as possible;
- $\alpha_{ij} \approx 1$ where the Galerkin solution is smooth.
- $\alpha_{ij} = \alpha_{ji}$.

Algebraic flux correction schemes

Equivalent system :

$$(\tilde{\mathbb{A}}\,\mathrm{U})_i = g_i + \sum_{j\neq i} \alpha_{ij}(U) f_{ij}$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j\neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.
The limiters α_{ij} should satisfy the following:

- $\alpha_{ij} \in [0,1]$;
- α_{ij} should be as close to 1 as possible;
- $\alpha_{ij} \approx 1$ where the Galerkin solution is smooth.
- $\alpha_{ij} = \alpha_{ji}$.

Algebraic flux correction schemes

Equivalent system :

$$(\mathbb{A}\,\mathrm{U})_i + \sum_{j \neq i} \big(1 - \alpha_{ij}(U)\big) f_{ij} = g_i$$

From the properties of $\mathbb{D}$ it follows that

$$\big(\mathbb{D}\mathrm{U}\big)_i = \sum_{j \neq i} f_{ij} \quad \text{where } f_{ij} = d_{ij}(u_j - u_i) \text{ are the fluxes}\,.$$

Goal : To limit the fluxes f_{ij} which are responsible for spurious oscillations.
The limiters α_{ij} should satisfy the following:

- $\alpha_{ij} \in [0, 1]$;
- α_{ij} should be as close to 1 as possible;
- $\alpha_{ij} \approx 1$ where the Galerkin solution is smooth.
- $\alpha_{ij} = \alpha_{ji}$.

University of Strathclyde Glasgow

The rewriting

A weak formulation : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot;\cdot,\cdot)$ is given by

$$d_h(u_h; u_h, v_h) =$$

The rewriting

A weak formulation : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot;\cdot,\cdot)$ is given by

$$d_h(u_h; u_h, v_h) = \sum_{i,j=1}^{N} (1 - \alpha_{ij}(u_h))\, d_{ij}(u_h(x_j) - u_h(x_i)) v_h(x_i)$$

The rewriting

A weak formulation : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot;\cdot,\cdot)$ is given by

$$d_h(u_h; u_h, v_h) = \text{...a couple of lines using the symmetry of } \alpha_{ij} \text{ and } d_{ij}\text{...}$$

The rewriting

A weak formulation : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot;\cdot,\cdot)$ is given by

$$d_h(u_h; u_h, v_h) = \sum_{E \in \mathscr{E}_h} (1 - \alpha_{ij}(u_h))\, |d_{ij}| h_E (\partial_{\mathbf{t}} u_h, \partial_{\mathbf{t}} v_h)_E\,.$$

The rewriting

A weak formulation : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \quad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot;\cdot,\cdot)$ is given by

$$d_h(u_h; u_h, v_h) = \sum_{E \in \mathscr{E}_h} (1 - \alpha_{ij}(u_h))\, |d_{ij}| h_E (\partial_{\mathbf{t}} u_h, \partial_{\mathbf{t}} v_h)_E\,.$$

Remark : Then, AFC schemes achieve a stable result by adding edge-based diffusion to the formulation.

University of Strathclyde Glasgow

The nonlinear edge-based diffusion

<u>The method</u> : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$\tilde{a}(u_h; v_h) := a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \qquad \forall\ v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot\,;\cdot,\ \cdot)$ is defined by

$$d_h(u_h; u_h, v_h) = \sum_{E \in \mathscr{E}_h} \gamma_0 \alpha_E(u_h) h_E^2 (\partial_{\mathbf{t}} u_h, \partial_{\mathbf{t}} v_h)_E\,.$$

Here, α_E are limiters defined as:

$$\alpha_E(w_h) := \max_{x \in E} \left[\xi_{w_h}(x)\right]^p\,,$$

with $\xi_{w_h} \in \mathbb{P}_1(\Omega)$ given by

$$\xi_{w_h}(x_i) := \begin{cases} \dfrac{\left|\sum_{j \in S_i} w_h(x_i) - w_h(x_j)\right|}{\sum_{j \in S_i} |w_h(x_i) - w_h(x_j)|}, & \text{if } \displaystyle\sum_{j \in S_i} |w_h(x_i) - w_h(x_j)| \neq 0\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

and $p \in \mathbb{N}$.

The nonlinear edge-based diffusion

The method : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$\tilde{a}(u_h; v_h) := a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \qquad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot\,;\cdot,\ \cdot)$ is defined by

$$d_h(u_h; u_h, v_h) = \sum_{E \in \mathscr{E}_h} \gamma_0 \alpha_E(u_h) h_E^2 (\partial_{\mathbf{t}} u_h, \partial_{\mathbf{t}} v_h)_E\,.$$

Here, α_E are limiters defined as:

$$\alpha_E(w_h) := \max_{x \in E} \left[\xi_{w_h}(x)\right]^p\,,$$

with $\xi_{w_h} \in \mathbb{P}_1(\Omega)$ given by

$$\xi_{w_h}(x_i) := \begin{cases} \dfrac{\left|\sum_{j \in S_i} w_h(x_i) - w_h(x_j)\right|}{\sum_{j \in S_i} |w_h(x_i) - w_h(x_j)|}, & \text{if } \displaystyle\sum_{j \in S_i} |w_h(x_i) - w_h(x_j)| \neq 0\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

and $p \in \mathbb{N}$.

The nonlinear edge-based diffusion

The method : Find $u_h \in \mathbb{P}_1(\Omega)$ such that

$$\tilde{a}(u_h; v_h) := a(u_h, v_h) + d_h(u_h; u_h, v_h) = (f, v_h)_\Omega \qquad \forall\, v_h \in \mathbb{P}_1(\Omega)\,.$$

The stabilisation term $d_h(\cdot\,;\cdot,\ \cdot)$ is defined by

$$d_h(u_h; u_h, v_h) = \sum_{E\in\mathscr{E}_h} \gamma_0 \alpha_E(u_h) h_E^2 (\partial_{\mathbf{t}} u_h, \partial_{\mathbf{t}} v_h)_E\,.$$

Here, α_E are limiters defined as:

$$\alpha_E(w_h) := \max_{x\in E} \left[\xi_{w_h}(x)\right]^p\ ,$$

with $\xi_{w_h} \in \mathbb{P}_1(\Omega)$ given by

$$\xi_{w_h}(x_i) := \begin{cases} \dfrac{\left|\sum_{j\in S_i} w_h(x_i) - w_h(x_j)\right|}{\sum_{j\in S_i} |w_h(x_i) - w_h(x_j)|}, & \text{if } \displaystyle\sum_{j\in S_i} |w_h(x_i) - w_h(x_j)| \neq 0\,, \\ 0, & \text{otherwise}\,, \end{cases}$$

and $p \in \mathbb{N}$.

Properties of the nonlinear diffusion term

- If the mesh is symmetric with respect to its interior nodes, then the method is linearity preserving, this is,

$$\alpha_E(z) = 0 \qquad \forall\, z \in \mathbb{P}_1(\omega_E)\,.$$

Theorem

The discrete problem has at least one solution. Moreover, if $C_{\text{lip}}\gamma_0\, h < \varepsilon$, where C_{lip} is the Lipschitz constant, then the solution is unique.

Properties of the nonlinear diffusion term

- If the mesh is symmetric with respect to its interior nodes, then the method is linearity preserving, this is,

$$\alpha_E(z) = 0 \qquad \forall\, z \in \mathbb{P}_1(\omega_E)\,.$$

- The nonlinear form $d_h(\cdot;\cdot,\cdot)$ is Lipschitz continuous. More precisely, there exists $C_{\text{lip}} > 0$, independent of h, such that, for all $v_h, w_h, z_h \in \mathbb{P}_1(\Omega)$, the following holds

$$|d_h(v_h; v_h, z_h) - d_h(w_h; w_h, z_h)| \le C_{\text{lip}}\gamma_0 h\, |v_h - w_h|_{1,\Omega}\, |z_h|_{1,\Omega}\,.$$

Theorem

The discrete problem has at least one solution. Moreover, if $C_{\text{lip}}\gamma_0\, h < \varepsilon$, where C_{lip} is the Lipschitz constant, then the solution is unique.

Properties of the nonlinear diffusion term

- If the mesh is symmetric with respect to its interior nodes, then the method is linearity preserving, this is,

$$\alpha_E(z) = 0 \qquad \forall\, z \in \mathbb{P}_1(\omega_E)\,.$$

- The nonlinear form $d_h(\cdot;\cdot,\cdot)$ is Lipschitz continuous. More precisely, there exists $C_{\rm lip} > 0$, independent of h, such that, for all $v_h, w_h, z_h \in \mathbb{P}_1(\Omega)$, the following holds

$$|d_h(v_h; v_h, z_h) - d_h(w_h; w_h, z_h)| \le C_{\rm lip}\gamma_0 h\, |v_h - w_h|_{1,\Omega}\, |z_h|_{1,\Omega}\,.$$

Theorem

The discrete problem has at least one solution. Moreover, if $C_{\rm lip}\gamma_0\, h < \varepsilon$, where $C_{\rm lip}$ is the Lipschitz constant, then the solution is unique.

The discrete maximum principle

Definition (DMP)

The semilinear form $\tilde{a}(\cdot;\cdot)$ is said to satisfy the *DMP property* if the following holds: If $u_h(x_i)$ is a local negative minimum, then there exist negative quantites $(c_E)_{E\in\mathscr{E}_i}$ such that

$$\tilde{a}(u_h;\psi_i) \leq \sum_{E\in\mathscr{E}_i} c_E \left|\partial_{\mathbf{t}} u_h|_E\right| .$$

The discrete maximum principle

Definition (DMP)

The semilinear form $\tilde{a}(\cdot;\cdot)$ is said to satisfy the *DMP property* if the following holds: If $u_h(x_i)$ is a local negative minimum, then there exist negative quantites $(c_E)_{E\in\mathscr{E}_i}$ such that

$$\tilde{a}(u_h;\psi_i) \leq \sum_{E\in\mathscr{E}_i} c_E \left|\partial_{\mathbf{t}} u_h|_E\right| .$$

Theorem

- *Let us suppose that the mesh is Delaunay. Then, if* $\gamma_0 > C_0\|\boldsymbol{b}\|_{\infty,E} + C_1\sigma h_E$*, the semilinear form* $\tilde{a}(\cdot;\cdot)$ *satisfies the DMP property.*

University of Strathclyde Glasgow

The discrete maximum principle

Definition (DMP)

The semilinear form $\tilde{a}(\cdot;\cdot)$ is said to satisfy the *DMP property* if the following holds: If $u_h(x_i)$ is a local negative minimum, then there exist negative quantites $(c_E)_{E\in\mathscr{E}_i}$ such that

$$\tilde{a}(u_h;\psi_i) \le \sum_{E\in\mathscr{E}_i} c_E \left|\partial_{\mathbf{t}} u_h|_E\right| .$$

Theorem

- *Let us suppose that the mesh is Delaunay. Then, if* $\gamma_0 > C_0\|\boldsymbol{b}\|_{\infty,E} + C_1\sigma h_E$, *the semilinear form* $\tilde{a}(\cdot;\cdot)$ *satisfies the DMP property.*
- *For an arbitrary regular mesh, the same result is valid if* $\gamma_0 > C_0\|\boldsymbol{b}\|_{\infty,E} + C_1\sigma h_E + C_2\varepsilon h_E^{-1}$.

University of Strathclyde Glasgow

Error estimates

Mesh-dependent norm:

$$\|v_h\|_h^2 := a(v_h, v_h) + d_h(u_h; v_h, v_h)\,.$$

General estimate: There exists $C > 0$, independent of ε such that

$$\|u - u_h\|_h \le C\,\|u - i_h u\|_{1,\Omega} + (d_h(u_h; i_h u, i_h u))^{\frac{1}{2}}\,.$$

Consistency error:

$$d_h(u_h; i_h u, i_h u) \le \gamma_0 \sum_{E \in \mathscr{E}_h} h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2 \le C\,h\,\|u\|_{1,\Omega}^2\,.$$

Error estimates

Mesh-dependent norm:

$$\|v_h\|_h^2 := a(v_h, v_h) + d_h(u_h; v_h, v_h)\,.$$

General estimate: There exists $C > 0$, independent of ε such that

$$\|u - u_h\|_h \le C\, \underbrace{\|u - i_h u\|_{1,\Omega}}_{\le C h |u|_{2,\Omega}} + (d_h(u_h; i_h u, i_h u))^{\frac{1}{2}}\,.$$

Consistency error:

$$d_h(u_h; i_h u, i_h u) \le \gamma_0 \sum_{E \in \mathscr{E}_h} h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2 \le C\, h\, \|u\|_{1,\Omega}^2\,.$$

Error estimates

Mesh-dependent norm:

$$\|v_h\|_h^2 := a(v_h, v_h) + d_h(u_h; v_h, v_h)\,.$$

General estimate: There exists $C > 0$, independent of ε such that

$$\|u - u_h\|_h \le C\, \underbrace{\|u - i_h u\|_{1,\Omega}}_{\le Ch|u|_{2,\Omega}} + \left(d_h(u_h; i_h u, i_h u)\right)^{\frac{1}{2}}\,.$$

Consistency error:

$$d_h(u_h; i_h u, i_h u) \le \gamma_0 \sum_{E \in \mathscr{E}_h} h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2 \le C\, h\, \|u\|_{1,\Omega}^2\,.$$

Theorem

There exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \le C h^{\frac{1}{2}} \|u\|_{2,\Omega}\,.$$

Error estimates

Mesh-dependent norm:

$$\|v_h\|_h^2 := a(v_h, v_h) + d_h(u_h; v_h, v_h)\,.$$

General estimate: There exists $C > 0$, independent of ε such that

$$\|u - u_h\|_h \leq C\ \underbrace{\|u - i_h u\|_{1,\Omega}}_{\leq Ch|u|_{2,\Omega}} + \left(d_h(u_h; i_h u, i_h u)\right)^{\frac{1}{2}}\,.$$

Consistency error:

$$d_h(u_h; i_h u, i_h u) \leq \gamma_0 \sum_{E \in \mathscr{E}_h} h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2 \leq C\,h\,\|u\|_{1,\Omega}^2\,.$$

Theorem

There exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \leq C h^{\frac{1}{2}} \|u\|_{2,\Omega}\,.$$

Error estimates

Mesh-dependent norm:

$$\|v_h\|_h^2 := a(v_h, v_h) + d_h(u_h; v_h, v_h)\,.$$

General estimate: There exists $C > 0$, independent of ε such that

$$\|u - u_h\|_h \le C\, \underbrace{\|u - i_h u\|_{1,\Omega}}_{\le Ch|u|_{2,\Omega}} + \left(d_h(u_h; i_h u, i_h u)\right)^{\frac{1}{2}}\,.$$

Consistency error:

$$d_h(u_h; i_h u, i_h u) \le \gamma_0 \sum_{E\in\mathscr{E}_h} h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2 \le C\, h\, \|u\|_{1,\Omega}^2\,.$$

Theorem

There exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \le C h^{\frac{1}{2}} \|u\|_{2,\Omega}\,.$$

An improved error estimate: The role of linearity preservation

$$d_h(i_h u; i_h u, i_h u) = \sum_{E \in \mathscr{E}_h} \gamma_0 \alpha_E(i_h u) h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2$$

Theorem

If the form $d_h(\cdot;\cdot,\cdot)$ is linearity preserving, then there exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \leq Ch \left(\|u\|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|u\|_{1,\Omega} \right) .$$

University of Strathclyde Glasgow

Error estimates

An improved error estimate: The role of linearity preservation

$$d_h(i_h u; i_h u, i_h u) = \sum_{E \in \mathscr{E}_h} \gamma_0 (\alpha_E(i_h u) - \alpha_E(i_E(u))) h_E^2 \|\partial_{\mathbf{t}} i_h u\|_{0,E}^2$$

Theorem

If the form $d_h(\cdot;\cdot,\cdot)$ is linearity preserving, then there exists a constant $C>0$, independent of h, such that

$$\|u - u_h\|_h \le C h \left(\|u\|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|u\|_{1,\Omega} \right).$$

University of Strathclyde Glasgow

An improved error estimate: The role of linearity preservation

$$d_h(i_h u; i_h u, i_h u) \leq Ch \left\{ \sum_{E \in \mathscr{E}_h} |i_h u - i_E u|^2_{1,\Omega_E} \right\}^{\frac{1}{2}} |i_h u|_{1,\Omega}$$

Theorem

If the form $d_h(\cdot;\cdot,\cdot)$ is linearity preserving, then there exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \leq Ch \left(\|u\|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|u\|_{1,\Omega} \right) .$$

University of Strathclyde Glasgow

An improved error estimate: The role of linearity preservation

$$d_h(i_h u; i_h u, i_h u) \le C h^2 |u|_{2,\Omega} |u|_{1,\Omega}\,.$$

Theorem

If the form $d_h(\cdot;\cdot,\cdot)$ is linearity preserving, then there exists a constant $C > 0$, independent of h, such that

$$\|u - u_h\|_h \le C h \left(\|u\|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|u\|_{1,\Omega} \right)\,.$$

Error estimates

An improved error estimate: The role of linearity preservation

$$d_h(i_h u; i_h u, i_h u) \le C h^2 |u|_{2,\Omega} |u|_{1,\Omega}\,.$$

Theorem

If the form $d_h(\cdot;\cdot,\cdot)$ is linearity preserving, then there exists a constant $C>0$, independent of h, such that

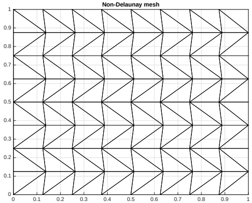
$$\|u - u_h\|_h \le C h \left(\|u\|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|u\|_{1,\Omega} \right)\,.$$

University of Strathclyde Glasgow

Numerical Results I : Convergence for smooth solutions

Data: $\varepsilon = 1$, $\sigma = 1$, $u(x, y) = \sin(2\pi x)\sin(2\pi y)$.

l	$\|e\|_{0,\Omega}$	ord.	$\|e\|_{1,\Omega}$	ord.	$\|e\|_h$	ord.
3	0.3835	–	3.5299	–	5.5746	–
4	0.1661	1.21	2.0053	0.82	2.4168	1.21
5	0.0451	1.88	0.9808	1.03	1.0317	1.23
6	0.0127	1.82	0.4811	1.03	0.4872	1.08
7	0.0042	1.59	0.2397	1.01	0.2405	1.02
8	0.0016	1.38	0.1198	1.00	0.1199	1.00



Data: $\varepsilon = 10^{-5}$, $\boldsymbol{b} = (-y, x)$, $\sigma = f = 0$.

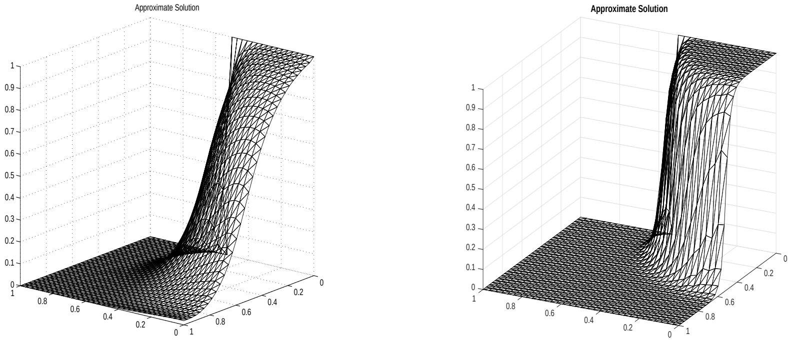


Figure 1 : Discrete solution for $p = 1$ (left) and $p = 8$ (right).

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$a(e,v) = \sum_{K\in\mathscr{T}} (f,v)_K - \varepsilon(\nabla u_h, \nabla v)_K - (\boldsymbol{b}\cdot\nabla u_h + \sigma u_h, v)_K$$

University of Strathclyde Glasgow

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$a(e,v) = \sum_{K \in \mathscr{T}} \left((\mathcal{R}_K, v)_K + \sum_{E \in \mathscr{E}_K \cap \mathscr{E}_I} (-\varepsilon \partial_{\boldsymbol{n}} u_h, v)_E \right)$$

University of Strathclyde Glasgow

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$a(e,v) = \sum_{K\in\mathscr{T}} \left((\mathcal{R}_K, v)_K + \sum_{E\in\mathscr{E}_K\cap\mathscr{E}_I} (-\varepsilon\partial_{\boldsymbol{n}} u_h, v)_E \right)$$

Equilibrated fluxes: We need a set of *boundary fluxes* $\{g_{E,K} : E \in \mathscr{E}_K, K \in \mathscr{T}\}$ satisfying:

Consistency:

$$g_{E,K} + g_{E,K'} = 0 \quad \text{if } E \in \mathscr{E}_K \cap \mathscr{E}_{K'},$$

Full first order equilibration: For all $n \in \mathcal{V}_K$ and all $K \in \mathscr{T}$,

$$0 = (f, \lambda_n)_K - a_K(u_h, \lambda_n) - d_h^K(u_h; u_h, \lambda_n) + \sum_{E\in\mathscr{E}_K} (g_{E,K}, \lambda_n)_E ,$$

for all basis function λ_n of $\mathbb{P}_1(\Omega)$.

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$a(e,v) = \sum_{K \in \mathscr{T}} \left((\mathcal{R}_K, v)_K + \sum_{E \in \mathscr{E}_K} (g_{E,K} - \varepsilon \partial_{\boldsymbol{n}} u_h, v)_E \right)$$

Equilibrated fluxes: We need a set of *boundary fluxes* $\{g_{E,K} : E \in \mathscr{E}_K, K \in \mathscr{T}\}$ satisfying:

Consistency:

$$g_{E,K} + g_{E,K'} = 0 \quad \text{if } E \in \mathscr{E}_K \cap \mathscr{E}_{K'},$$

Full first order equilibration: For all $n \in \mathcal{V}_K$ and all $K \in \mathscr{T}$,

$$0 = (f, \lambda_n)_K - a_K(u_h, \lambda_n) - d_h^K(u_h; u_h, \lambda_n) + \sum_{E \in \mathscr{E}_K} (g_{E,K}, \lambda_n)_E \ ,$$

for all basis function λ_n of $\mathbb{P}_1(\Omega)$.

University of Strathclyde Glasgow

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H^1_0(\Omega)$,

$$a(e,v) = \sum_{K\in\mathscr{T}}\Bigg((\underbrace{\mathcal{R}_K}_{=-\nabla\cdot\boldsymbol{\sigma}_K}, v)_K + \sum_{E\in\mathscr{E}_K} (\underbrace{g_{E,K} - \varepsilon\partial_{\boldsymbol{n}} u_h}_{=\boldsymbol{\sigma}_K\cdot\boldsymbol{n}}, v)_E \Bigg)$$

Equilibrated fluxes: We need a set of *boundary fluxes* $\{g_{E,K} : E \in \mathscr{E}_K, K \in \mathscr{T}\}$ satisfying:

Consistency:

$$g_{E,K} + g_{E,K'} = 0 \quad \text{if } E \in \mathscr{E}_K \cap \mathscr{E}_{K'},$$

Full first order equilibration: For all $n \in \mathcal{V}_K$ and all $K \in \mathscr{T}$,

$$0 = \; (f,\lambda_n)_K - a_K(u_h,\lambda_n) - d_h^K(u_h;u_h,\lambda_n) + \sum_{E\in\mathscr{E}_K} (g_{E,K},\lambda_n)_E \;,$$

for all basis function λ_n of $\mathbb{P}_1(\Omega)$.

University of Strathclyde Glasgow

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$a(e, v) = \sum_{K \in \mathscr{T}} (\boldsymbol{\sigma}_K, \nabla v)_K$$

University of Strathclyde Glasgow

A computable error bound

The residual equation : Defining $e = u - u_h$ one gets, for all $v \in H_0^1(\Omega)$,

$$\begin{aligned} a(e,v) &= \sum_{K\in\mathscr{T}} (\boldsymbol{\sigma}_K, \nabla v)_K \\ &\leq \sum_{K\in\mathscr{T}} \|\boldsymbol{\sigma}_K\|_{0,K} \|\nabla v\|_{0,K} \end{aligned}$$

The computable error bound

Theorem

For each element $K \in \mathscr{T}$, define a local error indicator by the rule

$$\eta_K = \frac{1}{\sqrt{\varepsilon}} \|\boldsymbol{\sigma}_K\|_{0,K} \,.$$

Then

$$\|e\|_\Omega^2 \leq \eta^2 := \sum_{K \in \mathscr{T}} \eta_K^2 .$$

Remark : $\|\boldsymbol{\sigma}_K\|_{0,K}$ can be computed analytically.

University of Strathclyde Glasgow

The computable error bound

Theorem

For each element $K \in \mathscr{T}$, define a local error indicator by the rule

$$\eta_K = \frac{1}{\sqrt{\varepsilon}} \|\boldsymbol{\sigma}_K\|_{0,K} .$$

Then

$$\|e\|_\Omega^2 \leq \eta^2 := \sum_{K \in \mathscr{T}} \eta_K^2 .$$

Remark : $\|\boldsymbol{\sigma}_K\|_{0,K}$ can be computed analytically.

University of Strathclyde Glasgow

The computable error bound: efficiency

Using known/familiar arguments :

$$\eta_K \leq C \sum_{K' \in \Omega_K} \left(\mathsf{C}_{K'} \|u - u_{\mathscr{T}}\|_{K'} + |d_h^{K'}(u_{\mathscr{T}}; u_{\mathscr{T}}, \lambda_n)| \right).$$

where

$$\mathsf{C}_K := \max \left\{ 1, Pe_K, \sqrt{\sigma} \frac{h_K}{\sqrt{\varepsilon}} \right\} .$$

Using the linearity preservation and Lipschitz continuity (as before):

$$d_h^{K'}(u_{\mathscr{T}}; u_{\mathscr{T}}, \lambda_n) \leq C \left(\frac{h_{K'}}{\varepsilon} \|u - u_{\mathscr{T}}\|_{\Omega_{K'}} + \frac{h_{K'}}{\sqrt{\varepsilon}} |u - \tilde{u}_{K'}|_{1,\Omega_{K'}} \right),$$

where $\tilde{u}_{K'}$ is the projection of u in $\mathbb{P}_1(\Omega_{K'})$.

The computable error bound: efficiency

Using known/familiar arguments :

$$\eta_K \leq C \sum_{K' \in \Omega_K} \left(\mathsf{C}_{K'} \|\!| u - u_{\mathscr{T}} \|\!|_{K'} + |d_h^{K'}(u_{\mathscr{T}}; u_{\mathscr{T}}, \lambda_n)| \right).$$

where

$$\mathsf{C}_K := \max \left\{ 1, Pe_K, \sqrt{\sigma} \frac{h_K}{\sqrt{\varepsilon}} \right\} .$$

Using the linearity preservation and Lipschitz continuity (as before):

$$d_h^{K'}(u_{\mathscr{T}}; u_{\mathscr{T}}, \lambda_n) \leq C \left(\frac{h_{K'}}{\varepsilon} \|\!| u - u_{\mathscr{T}} \|\!|_{\Omega_{K'}} + \frac{h_{K'}}{\sqrt{\varepsilon}} |u - \tilde{u}_{K'}|_{1,\Omega_{K'}} \right),$$

where $\tilde{u}_{K'}$ is the projection of u in $\mathbb{P}_1(\Omega_{K'})$.

The computable error bound: effciency

Theorem

There exists $C > 0$, independent of the size of the elements in the mesh $\mathscr{T}$, such that, for every $K \in \mathscr{T}$, the following local lower bound holds

$$\eta_K \le C \sum_{K' \in \Omega_K} \left(\left(\mathsf{C}_{K'} + \frac{h_{K'}}{\varepsilon} \right) |||u - u_{\mathscr{T}}|||_{K'} + \frac{h_{K'}}{\sqrt{\varepsilon}} |u - \tilde{u}_{K'}|_{1,K'} \right).$$

Numerical results

Data: $\varepsilon = 10^{-3}$, $\sigma = 1$, $u(x,y,z) = yz(1-y)(1-z)\left(x - \frac{e^{\frac{-(1-x)}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1-e^{-\frac{1}{\varepsilon}}}\right)$.

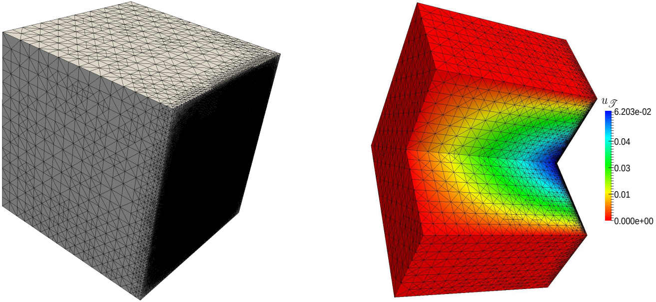


Figure 2 : 3D Example: Adapted mesh and isovalues of the solution.

Numerical results

Data: $\varepsilon = 10^{-3}$, $\sigma = 1$, $u(x,y,z) = yz(1-y)(1-z)\left(x - \dfrac{e^{\frac{-(1-x)}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1-e^{-\frac{1}{\varepsilon}}}\right)$.

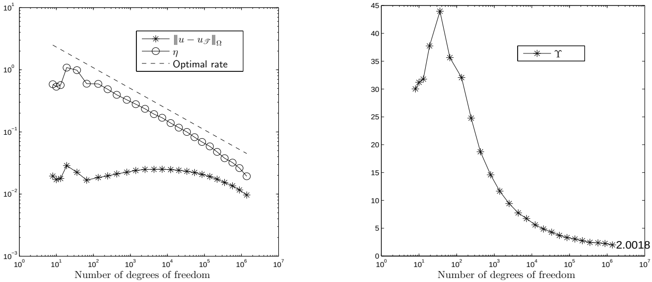


Figure 2 : 3D Example: Error estimator and effectivity indices.

Numerical results

Data: $\varepsilon = 10^{-4}$, $\sigma = 0$, $\boldsymbol{b} = (-y, x)$, $f = 0$.

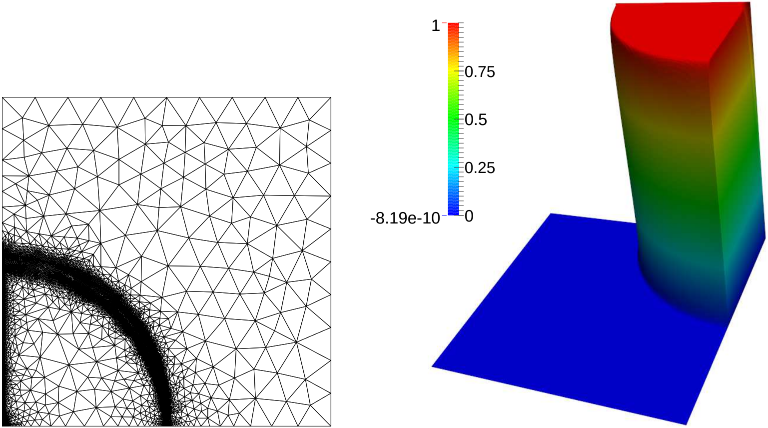


Figure 3 : Rotating converction field: adapted mesh, and discrete solution.

University of Strathclyde Glasgow

Conclusions and perspectives

1. AFC schemes can be rewritten as edge diffusion methods.
2. A new definition of the limiters: convergence in any regular mesh.
3. Linearity preservation and Lipschitz continuity provide:
    1. an improved error estimate;
    2. a computable error bound which can be proven to be locally efficient.

Future extensions:

- General meshes.
- Time-dependent problems.
- Nonlinear problems.
- ......

Conclusions and perspectives

1. AFC schemes can be rewritten as edge diffusion methods.
2. A new definition of the limiters: convergence in any regular mesh.
3. Linearity preservation and Lipschitz continuity provide:
    1. an improved error estimate;
    2. a computable error bound which can be proven to be locally efficient.

Future extensions:

- General meshes.
- Time-dependent problems.
- Nonlinear problems.
- ......