Introduction à l'optimisation

Stéphane Canu
stephane.canu@litislab.eu

ASI 4 - Optimisation sous contraintes et le lagrangien

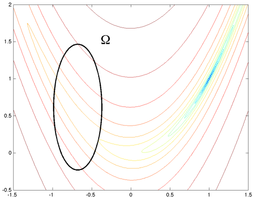
January 5, 2011

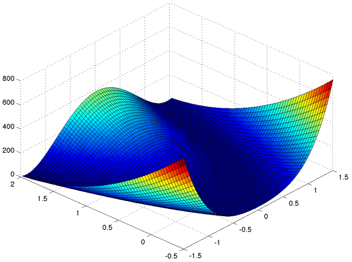
Plan

1. Optimisation sous contraintes et le lagrangien
   - Contraintes d'égalité
   - Conditions d'inégalité
   - Conclusion

Optimisation (convexe) sous contraintes

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) \\ \text{avec} & \mathbf{x} \in \Omega \quad \Leftrightarrow H(\mathbf{x}) = 0 \end{cases}$$



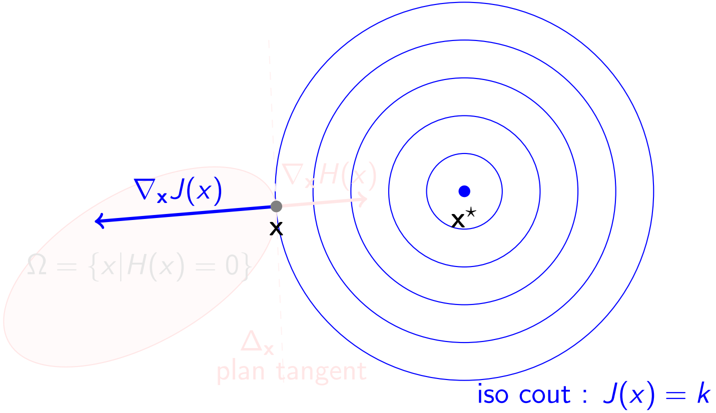


exemple :

$$\begin{cases} \min\limits_{x_1,x_2} & x_1 - 2x_2 \\ \text{avec} & x_1^2 + x_2^2 = 1 \end{cases}$$

Plan tangent à un ensemble de contraintes

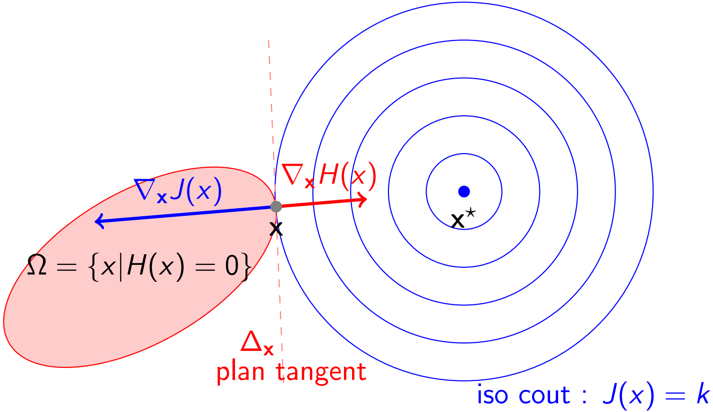
$$\begin{cases} \min\limits_{x_1,x_2} & J(\mathbf{x}) = (x_1-a)^2 - (x_2-b)^2 \\ \text{avec} & H(\mathbf{x}) = \alpha(x_1-c)^2 + \beta(x_2-d)^2 + \gamma x_1 x_2 - 1 \end{cases}$$



$$\nabla_{\mathbf{x}} H(x) = \lambda \, \nabla_{\mathbf{x}} J(x)$$

Plan tangent à un ensemble de contraintes

$$\begin{cases} \min\limits_{x_1,x_2} & J(\mathbf{x}) = (x_1 - a)^2 - (x_2 - b)^2 \\ \text{avec} & H(\mathbf{x}) = \alpha(x_1 - c)^2 + \beta(x_2 - d)^2 + \gamma x_1 x_2 - 1 \end{cases}$$



$$\nabla_{\mathbf{x}} H(x) = \lambda \, \nabla_{\mathbf{x}} J(x)$$

Le cas d'une seule contrainte d'égalité

$$\begin{cases} \min\limits_{\mathbf{x}} & J(\mathbf{x}) \\ \text{avec} & H(\mathbf{x}) = 0 \end{cases} \qquad \begin{array}{l} J(\mathbf{x}+\varepsilon\mathbf{d}) \approx J(\mathbf{x}) + \varepsilon\nabla_{\mathbf{x}} J(\mathbf{x})^\top \mathbf{d} \\ H(\mathbf{x}+\varepsilon\mathbf{d}) \approx H(\mathbf{x}) + \varepsilon\nabla_{\mathbf{x}} H(\mathbf{x})^\top \mathbf{d} \end{array}$$

cout J : $\mathbf{d}$ est une direction de descente, s'il existe un $\varepsilon_0 \in \mathbb{R}$ tel que $\forall \varepsilon \in \mathbb{R},\ 0 < \varepsilon \leq \varepsilon_0$

$$J(\mathbf{x}+\varepsilon\mathbf{d}) < J(\mathbf{x}) \qquad \Rightarrow \qquad \nabla_{\mathbf{x}} J(\mathbf{x})^\top \mathbf{d} < 0$$

contrainte H : $\mathbf{d}$ est une direction de descente admissible, s'il existe un $\varepsilon_0 \in \mathbb{R}$ tel que $\forall \varepsilon \in \mathbb{R},\ 0 < \varepsilon \leq \varepsilon_0$

$$H(\mathbf{x}+\varepsilon\mathbf{d}) = 0 \qquad \Rightarrow \qquad \nabla_{\mathbf{x}} H(\mathbf{x})^\top \mathbf{d} = 0$$

Si, au point $x^\star$, les vecteurs $\nabla_{\mathbf{x}} J(\mathbf{x}^\star)$ et $\nabla_{\mathbf{x}} H(\mathbf{x}^\star)$ sont colinéaires, il n'existe pas de direction $\mathbf{d}$ de descente admissible. $x^\star$ est donc une solution locale du problème.

Plan tangent à un ensemble de contraintes

L'ensemble des contraintes sur $\mathbb{R}^n$ s'écrit :

$$\left\{\begin{array}{l} H_1(\mathbf{x}) = 0 \\ H_2(\mathbf{x}) = 0 \\ \dots \\ H_p(\mathbf{x}) = 0 \end{array}\right.$$

Supposons que les fonctions H_i sont continûment différentiables.

Plan tangent aux contraintes en x

Si $\mathbf{x}$ est un point régulier[a] de la surface Ω définie par les contraintes $H_i(\mathbf{x}) = 0, i = 1, p$. Alors, le plan tangent aux contraintes est défini par :

$$\Delta_x = \left\{\mathbf{y} \in \mathbb{R}^n \;\middle|\; \nabla_x H_1(\mathbf{x})\mathbf{y} = 0,\ \nabla_x H_2(\mathbf{x})\mathbf{y} = 0, \dots,\ \nabla_x H_p(\mathbf{x})\mathbf{y} = 0\right\}$$

[a] les vecteurs gradients $\nabla_x H_i(\mathbf{x}), i = 1, p$ sont linéairement indépendants

Exemples de plan tangent

$p = 1,$

$$H_1(\mathbf{x}) = x_1^2 + x_2^2 - 1 \qquad \nabla_x H_1(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$$

$$\Delta_x = \left\{ \mathbf{y} \in \mathbb{R}^2 \mid 2x_1 y_1 + 2x_2 y_2 = 0 \right\}$$

$p = 2,$

$$\begin{cases} H_1(\mathbf{x}) = x_1^2 + x_2^2 + x_3^2 - 1 \\ H_2(\mathbf{x}) = x_1 - 2x_2 + 3x_3 - 1 \end{cases} \qquad \nabla_x H_1(\mathbf{x}) = \begin{pmatrix} 2x_1 \\ 2x_2 \\ 2x_3 \end{pmatrix}, \ \nabla_x H_2(\mathbf{x}) = \begin{pmatrix} 1 \\ -2 \\ 3 \end{pmatrix}$$

$$\Delta_x = \left\{ \mathbf{y} \in \mathbb{R}^3 \mid \begin{array}{l} 2x_1 y_1 + 2x_2 y_2 + 2x_3 y_3 = 0 \\ y_1 - 2y_2 + 3y_3 = 0 \end{array} \right\}$$

la droite tangente au cercle

conditions d'optimalité

J et H_i continûment différentiables : $\mathcal{P}\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) \\ \text{avec} & H_1(\mathbf{x}) = 0 \\ \text{et} & H_2(\mathbf{x}) = 0 \\ & \dots \\ & H_p(\mathbf{x}) = 0 \end{cases}$

Theorem (conditions d'optimalité)

Si $\mathbf{x}$ *est un point régulier par rapports aux contraintes* $H_i(\mathbf{x}) = 0, i = 1, p$ *et s'il est un extremum local de* J*,*
Alors, tout vecteur $\mathbf{y} \in \mathbb{R}^n$ *vérifiant*

$$\nabla_x H_1(\mathbf{x})\mathbf{y} = 0,\ \nabla_x H_2(\mathbf{x})\mathbf{y} = 0, \dots,\ \nabla_x H_p(\mathbf{x})\mathbf{y} = 0$$

doit aussi vérifier

$$\nabla_x J(\mathbf{x})\mathbf{y} = 0$$

démonstration

Multiplicateurs de Lagrange

supposons que J et les fonctions H_i sont continûment différentiables.

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) & \\ \text{avec} & H_1(\mathbf{x}) = 0 & \lambda_1 \\ \text{et} & H_2(\mathbf{x}) = 0 & \lambda_2 \\ & \dots & \\ & H_p(\mathbf{x}) = 0 & \lambda_p \end{cases}$$

à chaque contrainte est associé un réel λ_i : le multiplicateur de Lagrange.

Theorem (conditions d'optimalité du premier ordre)

Pour qu'un point $\mathbf{x}$ *soit un extremum local le* $\mathcal{P}$, *il est nécessaire que :*

$$\nabla_x J(x) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \qquad et \qquad H_i(\mathbf{x}) = 0, \quad i = 1, p$$

<u>démonstration</u>

$n+p$ inconnues ($\mathbf{x}$ et λ) et donc $n+p$ équations. Chaque contrainte ajoute une équation ($H_i(\mathbf{x}) = 0$) et une variable (λ_i)

Multiplicateurs de Lagrange

supposons que J et les fonctions H_i sont continûment différentiables.

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) & \\ \text{avec} & H_1(\mathbf{x}) = 0 & \color{red}{\lambda_1} \\ \text{et} & H_2(\mathbf{x}) = 0 & \color{red}{\lambda_2} \\ & \dots & \\ & H_p(\mathbf{x}) = 0 & \color{red}{\lambda_p} \end{cases}$$

à chaque contrainte est associé un réel λ_i : le multiplicateur de Lagrange.

Theorem (conditions d'optimalité du premier ordre)

Pour qu'un point $\mathbf{x}$ *soit un extremum local le* $\mathcal{P}$*, il est nécessaire que :*

$$\nabla_x J(x) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \qquad et \qquad H_i(\mathbf{x}) = 0, \quad i = 1, p$$

démonstration

$n+p$ inconnues ($\mathbf{x}$ et λ) et donc $n+p$ équations. Chaque contrainte ajoute une équation ($H_i(\mathbf{x}) = 0$) et une variable (λ_i)

Multiplicateurs de Lagrange

supposons que J et les fonctions H_i sont continûment différentiables.

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) & \\ \text{avec} & H_1(\mathbf{x}) = 0 & \lambda_1 \\ \text{et} & H_2(\mathbf{x}) = 0 & \lambda_2 \\ & \dots & \\ & H_p(\mathbf{x}) = 0 & \lambda_p \end{cases}$$

à chaque contrainte est associé un réel λ_i : le multiplicateur de Lagrange.

Theorem (conditions d'optimalité du premier ordre)

Pour qu'un point $\mathbf{x}$ *soit un extremum local le* $\mathcal{P}$*, il est nécessaire que :*

$$\nabla_x J(x) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \qquad \textit{et} \qquad H_i(\mathbf{x}) = 0, \;\; i = 1, p$$

<u>démonstration</u>

$n+p$ inconnues ($\mathbf{x}$ et λ) et donc $n+p$ équations. Chaque contrainte ajoute une équation ($H_i(\mathbf{x}) = 0$) et une variable (λ_i)

Un exemple où ça marche

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^3} & J(\mathbf{x}) = -x_1x_2 - x_1x_3 - x_2x_3 \\ \text{avec} & H(\mathbf{x}) = x_1 + x_2 + x_3 - 3 = 0 \end{cases}$$

$$\nabla_x J(x) = -\begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix} \qquad \nabla_x H(x) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Les conditions d'optimalité sont

$$\begin{cases} & -x_2 & -x_3 & +\lambda & = 0 \\ -x_1 & & -x_3 & +\lambda & = 0 \\ -x_1 & -x_2 & & +\lambda & = 0 \\ x_1 & +x_2 & +x_3 & & = 3 \end{cases}$$

la résolution du système $(A\backslash b)$ donne :

$$x_1 = x_2 = x_3 = 1 \;\text{ et }\; \lambda = 2$$

Un exemple où ça marche

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^3} & J(\mathbf{x}) = -x_1x_2 - x_1x_3 - x_2x_3 \\ \text{avec} & H(\mathbf{x}) = x_1 + x_2 + x_3 - 3 = 0 \end{cases}$$

$$\nabla_x J(x) = -\begin{pmatrix} x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix} \qquad \nabla_x H(x) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Les conditions d'optimalité sont

$$\begin{cases} & -x_2 & -x_3 & +\lambda & = 0 \\ -x_1 & & -x_3 & +\lambda & = 0 \\ -x_1 & -x_2 & & +\lambda & = 0 \\ x_1 & +x_2 & +x_3 & & = 3 \end{cases}$$

la résolution du système $(A\backslash b)$ donne :

$$x_1 = x_2 = x_3 = 1 \ \text{ et } \lambda = 2$$

Un autre exemple où ça marche

$$\begin{cases} \displaystyle\min_{\mathbf{x}\in\mathbb{R}^3} & J(\mathbf{x}) = x_1^2 - 4x_1 - x_1x_2 + x_1x_3 + x_2x_3 \\ \text{avec} & H(\mathbf{x}) = x_1 + x_2 + x_3 - 1 = 0 \end{cases}$$

$$\nabla_x J(x) = \begin{pmatrix} 2x_1 - 4 - x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix} \qquad \nabla_x H(x) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Les conditions d'optimalité sont

$$\begin{cases} 2x_1 & -x_2 & +x_3 & +\lambda & = 4 \\ -x_1 & & +x_3 & +\lambda & = 0 \\ x_1 & +x_2 & & +\lambda & = 0 \\ x_1 & +x_2 & +x_3 & & = 1 \end{cases}$$

la résolution du système $(A\backslash b)$ donne :

$$x_1 = 1, \;\; x_2 = -1, \;\; x_3 = 1 \;\; \text{et } \lambda = 0$$

$\lambda = 0$: la contrainte est inutile (inactive (ou passive) et non saturée)

Un autre exemple où ça marche

$$\begin{cases} \displaystyle\min_{\mathbf{x}\in\mathbb{R}^3} & J(\mathbf{x}) = x_1^2 - 4x_1 - x_1x_2 + x_1x_3 + x_2x_3 \\ \text{avec} & H(\mathbf{x}) = x_1 + x_2 + x_3 - 1 = 0 \end{cases}$$

$$\nabla_x J(x) = \begin{pmatrix} 2x_1 - 4 - x_2 + x_3 \\ x_1 + x_3 \\ x_1 + x_2 \end{pmatrix} \qquad \nabla_x H(x) = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Les conditions d'optimalité sont

$$\begin{cases} 2x_1 & -x_2 & +x_3 & +\lambda & = 4 \\ -x_1 & & +x_3 & +\lambda & = 0 \\ x_1 & +x_2 & & +\lambda & = 0 \\ x_1 & +x_2 & +x_3 & & = 1 \end{cases}$$

la résolution du système $(A\backslash b)$ donne :

$$x_1 = 1, \quad x_2 = -1, \quad x_3 = 1 \quad \text{et } \lambda = 0$$

$\lambda = 0$: la contrainte est inutile (inactive (ou passive) et non saturée)

Un autre exemple où ça marche

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) = -\frac{1}{2}\,\mathbf{x}^\top A\mathbf{x} - \mathbf{x}^\top\mathbf{b} \\ \text{avec} & C\mathbf{x} = \mathbf{d} \end{cases}$$

$$\begin{aligned} \nabla_{\mathbf{x}} J &= A\mathbf{x} - \mathbf{b} \\ \nabla_{\mathbf{x}} H &= C^\top \end{aligned}$$

$$\begin{aligned} \nabla_x J(x) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \quad &\Rightarrow \quad A\mathbf{x} + C^\top\lambda &= \mathbf{b} \\ H(\mathbf{x}) = 0 \quad &\Rightarrow \quad C\mathbf{x} &= \mathbf{d} \end{aligned}$$

...et on résoud le système linéaire.

Attention si en plus on cherche $\mathbf{x} \geq 0$ ça devient plus compliqué

Un exemple où ça ne marche pas

$$\begin{cases} \min\limits_{x_1, x_2} & x_1 + x_2 \\ \text{avec} & (x_1+1)^2 + x_2^2 = 1 \\ \text{et} & (x_1-2)^2 + x_2^2 = 4 \end{cases}$$

le minimum est $(0, 0)$ l'unique solution réalisable ! Dans ce cas, il n'existe pas de multiplicateurs de Lagrange

lagrangien

Une fonction bien pratique :

définition : lagrangien

On appelle lagrangien du problème $\mathcal{P}$ la fonction L définie par :

$$L(\mathbf{x}, \lambda) = J(x) + \sum_{i=1}^{p} \lambda_i H_i(\mathbf{x})$$

Grâce au lagrangien on retrouve les conditions d'optimalité :

$$\begin{cases} \nabla_x L(\mathbf{x}, \lambda) = 0 & \Rightarrow \quad \nabla_x J(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \\ \nabla_{\lambda_i} L(\mathbf{x}, \lambda) = 0 & \Rightarrow \quad H_i(\mathbf{x}) = 0 \end{cases}$$

Interprétation graphique : optimisation multicritères.

Sensibilité et perturbation

$$\mathcal{P}_\varepsilon = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) \\ \text{avec} & H(\mathbf{x}) = \varepsilon \end{cases} \qquad\qquad \varepsilon \in \mathbb{R}^p$$

Theorem (Sensibilité)

$$\partial_{\varepsilon_i} J\big(\mathbf{x}(\varepsilon)\big)\Big|_{\varepsilon=0} = -\lambda_i \qquad\qquad i = 1, p$$

démonstration

$$\begin{aligned} \nabla_\varepsilon J\big(x(\varepsilon)\big)\Big|_{\varepsilon=0} &= \nabla_\varepsilon \mathbf{x}(0)\ \nabla_{\mathbf{x}} J\big(\mathbf{x}(0)\big) \\ \nabla_\varepsilon H\big(x(\varepsilon)\big)\Big|_{\varepsilon=0} &= \nabla_\varepsilon \mathbf{x}(0)\ \nabla_{\mathbf{x}} H\big(\mathbf{x}(0)\big) \\ \nabla_\varepsilon H\big(x(\varepsilon)\big) &= I \end{aligned}$$

or

$$\nabla_x J(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i \nabla_x H_i(\mathbf{x}) = 0 \quad \Rightarrow \quad \underbrace{\nabla_\varepsilon \mathbf{x}(0)\ \nabla_x J(\mathbf{x})}_{\nabla_\varepsilon J\left(x(\varepsilon)\right)\Big|_{\varepsilon=0}} + \underbrace{\nabla_\varepsilon \mathbf{x}(0)\ \nabla_x H(\mathbf{x})}_{I} \lambda = 0$$

Le multiplicateur de Lagrange indique une forme de l'influence de la contrainte dans le cout de la solution

Sensibilité : preuve pour une contrainte

$$x^* = \underset{x \in \mathbb{R}^n}{\operatorname{Arg\,min}} \; J(x) \qquad H(x) = 0$$

$$x^*(\varepsilon) = \underset{x \in \mathbb{R}^n}{\operatorname{Arg\,min}} \; J(x) \qquad H(x) = \varepsilon$$

$H : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$J'_\varepsilon(x) = [\nabla_x J(x)]^T \cdot \nabla_\varepsilon x = \sum_{i=1}^{n} \frac{\partial J}{\partial x_i} \cdot \frac{\partial x_i}{\partial \varepsilon}$$

$$H'_\varepsilon(x) = 1$$

$$H'_\varepsilon(x) = [\nabla_x H(x)]^T \nabla_\varepsilon(x) = \sum_{i=1}^{n} \frac{\partial H}{\partial x_i} \cdot \frac{\partial x_i}{\partial \varepsilon}$$

$$\nabla L = 0 \iff \nabla_x J(x^*) + \lambda \nabla_x H(x^*) = 0$$

$$\underbrace{\nabla_\varepsilon x^T \nabla_x J(x^*)}_{J'_\varepsilon(x^*)} + \lambda \underbrace{\nabla_\varepsilon x^T \nabla_x H(x^*)}_{= H'_\varepsilon(x^*) = 1} = 0$$

$$\boxed{J'_\varepsilon(x^*) = -\lambda}$$

Conditions du second ordre

Theorem (Conditions du second ordre)

Si un point $\mathbf{x}$ *vérifie les conditions d'optimalité du premier ordre et si la matrice*

$$M = \nabla_x^2 J(\mathbf{x}) + \sum_{i=1}^{p} \lambda_i \nabla_x^2 H_i(\mathbf{x})$$

est définie positive sur le plan tangent $\Delta_{\mathbf{x}}$*,*
Alors $\mathbf{x}$ *est une solution locale de* $\mathcal{P}$*.*

Exemple :

$$\mathcal{P}_1 = \begin{cases} \min\limits_{\mathbf{x} \in \mathbb{R}^3} & J(\mathbf{x}) = -x_1 x_2 - x_1 x_3 - x_2 x_3 \\ \text{avec} & H(\mathbf{x}) = x_1 + x_2 + x_3 - 3 = 0 \end{cases} \qquad M_1 = \begin{pmatrix} 0 & -1 & -1 \\ -1 & 0 & -1 \\ -1 & -1 & 0 \end{pmatrix}$$

sur le plan tangent $\Delta_{\mathbf{x}}$ on a $y_1 + y_2 + y_3 = 0$ et donc

$$\mathbf{y}^\top M \mathbf{y} = y_1^2 + y_2^2 + y_3^2 \geq 0$$

Plan

1. **Optimisation sous contraintes et le lagrangien**
   - Contraintes d'égalité
   - Conditions d'inégalité
   - Conclusion

Le cas d'une seule contrainte d'inégalité

$$\begin{cases} \min_{\mathbf{x}} & J(\mathbf{x}) \\ \text{avec} & H(\mathbf{x}) \leq 0 \end{cases} \qquad \begin{aligned} J(\mathbf{x}+\varepsilon\mathbf{d}) &\approx J(\mathbf{x}) + \varepsilon \nabla_{\mathbf{x}} J(\mathbf{x})^\top \mathbf{d} \\ H(\mathbf{x}+\varepsilon\mathbf{d}) &\approx H(\mathbf{x}) + \varepsilon \nabla_{\mathbf{x}} H(\mathbf{x})^\top \mathbf{d} \end{aligned}$$

cout J : $\mathbf{d}$ est une direction de descente, s'il existe un $\varepsilon_0 \in \mathbb{R}$ tel que $\forall \varepsilon \in \mathbb{R},\ 0 < \varepsilon \leq \varepsilon_0$

$$J(\mathbf{x}+\varepsilon\mathbf{d}) < J(\mathbf{x}) \qquad \Rightarrow \qquad \nabla_{\mathbf{x}} J(\mathbf{x})^\top \mathbf{d} < 0$$

contrainte H : $\mathbf{d}$ est une direction de descente admissible, s'il existe un $\varepsilon_0 \in \mathbb{R}$ tel que $\forall \varepsilon \in \mathbb{R},\ 0 < \varepsilon \leq \varepsilon_0$

$$H(\mathbf{x}+\varepsilon\mathbf{d}) \leq 0 \qquad \Rightarrow \qquad \begin{aligned} H(\mathbf{x}) < 0 : &\quad \mathbf{d} \text{ est quelconque} \\ H(\mathbf{x}) = 0 : &\quad \nabla_{\mathbf{x}} H(\mathbf{x})^\top \mathbf{d} < 0 \end{aligned}$$

Si, au point $x^\star$, on se trouve au bord du domaine admissible ($H(\mathbf{x}^\star) = 0$) et si les vecteurs $\nabla_{\mathbf{x}} J(\mathbf{x}^\star)$ et $\nabla_{\mathbf{x}} H(\mathbf{x}^\star)$ sont colinéaires et dans le sens opposé, il n'existe pas de direction $\mathbf{d}$ de descente admissible. $x^\star$ est donc une solution locale du problème.

Contraintes d'inégalité

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) \\ \text{avec} & H(\mathbf{x}) = 0 \\ \text{et} & G(\mathbf{x}) \leq 0 \end{cases}$$

definition : conditions de Karush, Kuhn et Tucker

stationarité $\nabla J(x^\star) + \sum_{j=1}^{p} \lambda_j \nabla h_j(x^\star) + \sum_{i=1}^{q} \mu_i \nabla g_i(x^\star) = 0$

admissibilité primal $g_i(x^\star) \leq 0 \qquad i = 1, \dots, q$

$h_j(x^\star) = 0 \qquad j = 1, \dots, p$

admissibilité duale $\mu_i \geq 0 \qquad i = 1, \dots, q$

complementarité $\mu_i g_i(x^\star) = 0 \qquad i = 1, \dots, q$

On les appelle aussi les *KKT conditions*

Conditions d'optimalité du premier ordre

Theorem

Sous certaines conditions de régularité, une solution $\mathbf{x}^\star$ *du problème d'optimisation vérifie nécessairement les conditions de Karush, Kuhn et Tucker*

Exemple

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^2} & J(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{avec} & 2x_1 + x_2 \leq -4 \end{cases}$$

stationarité $\begin{aligned} 2x_1 + 2\mu &= 0 \\ 2x_2 + \mu &= 0 \end{aligned}$

admissibilité primal $2x_1 + x_2 + 4 \leq 0$

admissibilité duale $\mu \geq 0$

complementarité $\mu(2x_1 + x_2 + 4) = 0$

$$x_1 = \frac{8}{5},\ x_2 = \frac{4}{5},\ \mu = \frac{8}{5},$$

Exemple

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^2} & J(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{avec} & 2x_1 + x_2 \leq -4 \end{cases}$$

stationarité $\begin{aligned} 2x_1 + 2\mu &= 0 \\ 2x_2 + \mu &= 0 \end{aligned}$

admissibilité primal $2x_1 + x_2 + 4 \leq 0$

admissibilité duale $\mu \geq 0$

complementarité $\mu(2x_1 + x_2 + 4) = 0$

$$x_1 = \frac{8}{5},\ x_2 = \frac{4}{5},\ \mu = \frac{8}{5},$$

Lagrangien

$$\mathcal{P} = \begin{cases} \min_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) \\ \text{avec} & H(\mathbf{x}) = 0 \\ \text{et} & G(\mathbf{x}) \leq 0 \end{cases}$$

Une fonction bien pratique :

définition : lagrangien

On appelle lagrangien du problème $\mathcal{P}$ la fonction L définie par :

$$L(\mathbf{x}, \lambda, \mu) = J(x) + \sum_{i=1}^{p} \lambda_i H_i(\mathbf{x}) + \sum_{i=1}^{q} \mu_i G_i(\mathbf{x})$$

définition : la fonction duale lagrangienne

On appelle fonction duale lagrangienne du problème $\mathcal{P}$ la fonction $\mathcal{L}$:

$$\mathcal{L}(\lambda, \mu) = \min_{\mathbf{x}\in\mathbb{R}^n} L(\mathbf{x}, \lambda, \mu)$$

Exemple de dualité

$$\mathcal{P} = \begin{cases} \min_{\mathbf{x} \in \mathbb{R}^2} & J(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{avec} & 2x_1 + x_2 \leq -4 \end{cases}$$

$$L(\mathbf{x}, \mu) = x_1^2 + x_2^2 + \mu(2x_1 + x_2 + 4)$$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mu) = 0 \;\Leftrightarrow\; \begin{array}{ll} 2x_1 + 2\mu & = 0 \\ 2x_2 + \mu & = 0 \end{array}$$

$$\mathcal{L}(\mu) = -\frac{5}{4}\mu^2 + 4\mu$$

$$\nabla_{\mu} \mathcal{L}(\mu) = 0 \;\Leftrightarrow\; \mu = \frac{8}{5}$$

Exemple de dualité

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^2} & J(\mathbf{x}) = x_1^2 + x_2^2 \\ \text{avec} & 2x_1 + x_2 \leq -4 \end{cases}$$

$$L(\mathbf{x}, \mu) = x_1^2 + x_2^2 + \mu(2x_1 + x_2 + 4)$$

$$\nabla_{\mathbf{x}} L(\mathbf{x}, \mu) = 0 \;\Leftrightarrow\; \begin{array}{ll} 2x_1 + 2\mu & = 0 \\ 2x_2 + \mu & = 0 \end{array}$$

$$\mathcal{L}(\mu) = -\frac{5}{4}\mu^2 + 4\mu$$

$$\nabla_{\mu}\mathcal{L}(\mu) = 0 \;\Leftrightarrow\; \mu = \frac{8}{5}$$

Exemple de dualité

$$\mathcal{P} = \begin{cases} \min_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} \\ \text{avec} & A\mathbf{x} = \mathbf{b} \\ \text{et} & \mathbf{x} \geq 0 \end{cases}$$

$$L(\mathbf{x}, \lambda, \mu) = \mathbf{c}^\top \mathbf{x} + \lambda^\top (A\mathbf{x} - \mathbf{b}) - \sum_{i=1}^{n} \mu_i x_i \qquad \mu \geq 0$$

le dual lagrangien

$$\begin{aligned} \mathcal{L}(\lambda, \mu) &= -\lambda^\top \mathbf{b} + \min_{\mathbf{x}} (\mathbf{c} + A^\top \lambda - \mu)^\top \mathbf{x} \\ &= \begin{cases} -\mathbf{b}^\top \lambda & \text{si } \mathbf{c} + A^\top \lambda - \mu = 0 \\ -\infty & \text{sinon} \end{cases} \end{aligned}$$

formulation duale :

$$\mathcal{D} = \begin{cases} \max_{\lambda\in\mathbb{R}^p} & \mathbf{b}^\top \lambda \\ \text{avec} & A^\top \lambda + \mathbf{c} \geq 0 \end{cases}$$

Exemple de dualité

$$\mathcal{P} = \begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) = \mathbf{c}^\top \mathbf{x} \\ \text{avec} & A\mathbf{x} = \mathbf{b} \\ \text{et} & \mathbf{x} \geq 0 \end{cases}$$

$$L(\mathbf{x}, \lambda, \mu) = \mathbf{c}^\top \mathbf{x} + \lambda^\top (A\mathbf{x} - \mathbf{b}) - \sum_{i=1}^{n} \mu_i x_i \qquad \mu \geq 0$$

le dual lagrangien

$$\begin{aligned} \mathcal{L}(\lambda, \mu) &= -\lambda^\top \mathbf{b} + \min_{\mathbf{x}} \big(\mathbf{c} + A^\top \lambda - \mu\big)^\top \mathbf{x} \\ &= \begin{cases} -\mathbf{b}^\top \lambda & \text{si } \mathbf{c} + A^\top \lambda - \mu = 0 \\ -\infty & \text{sinon} \end{cases} \end{aligned}$$

formulation duale :

$$\mathcal{D} = \begin{cases} \max\limits_{\lambda\in\mathbb{R}^p} & \mathbf{b}^\top \lambda \\ \text{avec} & A^\top \lambda + \mathbf{c} \geq 0 \end{cases}$$

Un autre exemple où ça marche

$$\begin{cases} \min\limits_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^\top A\mathbf{x} - \mathbf{x}^\top \mathbf{b} \\ \text{avec} & C\mathbf{x} = \mathbf{d} \\ \text{et} & \mathbf{x} \geq 0 \end{cases}$$

$$L(\mathbf{x},\lambda,\mu) = \frac{1}{2}\,\mathbf{x}^\top A\mathbf{x} - \mathbf{x}^\top \mathbf{b} + \lambda^\top (C\mathbf{x} - \mathbf{d}) - \sum_{i=1}^{n} \mu_i x_i \qquad \mu \geq 0$$

Pour calculer le dual lagrangien nous écrivons les conditions d'optimalité :

$$\nabla_{\mathbf{x}} L(\mathbf{x},\lambda,\mu) = 0 \quad \Rightarrow \quad A\mathbf{x} - \mathbf{b} + C^\top \lambda - \mu = 0$$

au minimum on a nécessairement $\mathbf{x} = A^{-1}(\mathbf{b} - C^\top \lambda + \mu)$. On utilise cette relation pour éliminer $\mathbf{x}$ du lagrangien et construire ainsi le dual lagangien

Un autre exemple où ça marche

$$\begin{cases} \min_{\mathbf{x}\in\mathbb{R}^n} & J(\mathbf{x}) = \frac{1}{2}\,\mathbf{x}^\top A\mathbf{x} - \mathbf{x}^\top \mathbf{b} \\ \text{avec} & C\mathbf{x} = \mathbf{d} \\ \text{et} & \mathbf{x} \geq 0 \end{cases}$$

$$L(\mathbf{x},\lambda,\mu) = \frac{1}{2}\,\mathbf{x}^\top A\mathbf{x} - \mathbf{x}^\top \mathbf{b} + \lambda^\top (C\mathbf{x} - \mathbf{d}) - \sum_{i=1}^{n} \mu_i x_i \qquad \mu \geq 0$$

Pour calculer le dual lagrangien nous écrivons les conditions d'optimalité :

$$\nabla_{\mathbf{x}} L(\mathbf{x},\lambda,\mu) = 0 \quad \Rightarrow \quad A\mathbf{x} - \mathbf{b} + C^\top\lambda - \mu = 0$$

au minimum on a nécessairement $\mathbf{x} = A^{-1}(\mathbf{b} - C^\top\lambda + \mu)$. On utilise cette relation pour éliminer $\mathbf{x}$ du lagrangien et construire ainsi le dual lagangien

Compléments

- conditions du second ordre
- stabilité - perturbations
- primal/dual : pour tout point $\mathbf{x}$ admissible (vérifiant les contraintes)

$$\text{cout dual} \quad \mathcal{L}(\lambda, \mu) \leq L(\mathbf{x}, \lambda, \mu) \leq J(\mathbf{x}) \quad \text{cout primal}$$

Plan

1. Optimisation sous contraintes et le lagrangien
   - Contraintes d'égalité
   - Conditions d'inégalité
   - Conclusion

Conclusion

- comment éliminer les contraintes ?
  - ajouter une variable par contrainte
- variable = multiplicateurs de Lagrange (l'influence de la contrainte)
  - contrainte d'égalité : multiplicateur quelconque
  - contrainte d'inégalité : multiplicateur positif (attention au sens)
- conditions KKT : optimalité (notamment dans le cas convexe)
- lagrangien : formulation duale

$$J(x) + \sum_{i=1}^{p} I_0\big(H_i(\mathbf{x})\big) + \sum_{i=1}^{q} I_-\big(G_i(\mathbf{x})\big)$$

$$L(\mathbf{x}, \lambda, \mu) = J(x) + \sum_{i=1}^{p} \lambda_i H_i(\mathbf{x}) + \sum_{i=1}^{q} \mu_i G_i(\mathbf{x})$$

  - I_0 la fonction indicatrice de zéro
  - $I_- = 0$ si $\mathbf{x} \leq 0$ et ∞ sinon

Conclusion

- comment éliminer les contraintes ?
  - ajouter une variable par contrainte

- variable = multiplicateurs de Lagrange (l'influence de la contrainte)
  - contrainte d'égalité : multiplicateur quelconque
  - contrainte d'inégalité : multiplicateur positif (attention au sens)

- conditions KKT : optimalité (notamment dans le cas convexe)

- lagrangien : formulation duale

$$L(\mathbf{x}, \lambda, \mu) = \begin{array}{lll} J(x) + \sum_{i=1}^{p} I_0\big(H_i(\mathbf{x})\big) & + \sum_{i=1}^{q} I_-\big(G_i(\mathbf{x})\big) \\ J(x) + \sum_{i=1}^{p} \lambda_i H_i(\mathbf{x}) & + \sum_{i=1}^{q} \mu_i G_i(\mathbf{x}) \end{array}$$

  - I_0 la fonction indicatrice de zéro
  - $I_- = 0$ si $\mathbf{x} \leq 0$ et ∞ sinon