Module B5

原始-对偶 分解法 Éclatement primal-dual

Sauf mention contraire, $\mathcal{X}$ (resp. $\mathcal{Y}$) est un espace de Hilbert, muni du produit scalaire noté $\langle\cdot,\cdot\rangle$ et on note $\|\cdot\|$ la norme qui découle du produit scalaire.

Dans ce module, on s'intéresse au problème de minimisation suivant

$$\min_{x\in\mathcal{X}} J(x) \tag{$\mathcal{P}$}$$

où $J : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ est une fonction s.c.i. de domaine non vide. On supposera que ce problème admet au moins une solution.

1 Principe et observations

L'idée dans l'utilisation d'un éclatement primal-dual d'opérateurs pour résoudre le problème ($\mathcal{P}$) est d'introduire une fonction de couplage $\mathcal{L} : \mathcal{X}\times\mathcal{Y} \to \overline{\mathbb{R}}$ vérifiant
耦合函数

$$\forall\, x\in\mathcal{X},\qquad J(x) = \sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

de sorte que, par dualité min-max (voir module **A6 : Dualité min-max**), le problème de minimisation (dit primal) ($\mathcal{P}$) s'écrive comme le problème de recherche de point-selle

$$\min_{x\in\mathcal{X}}\sup_{y\in\mathcal{Y}} \mathcal{L}(x;y)$$

On associe à ce problème un problème dit dual :

$$\max_{y\in\mathcal{Y}} E(y) \qquad \text{avec} \qquad E(y) = \inf_{x\in\mathcal{X}} \mathcal{L}(x;y) \tag{$\mathcal{D}$}$$

Si le problème primal ($\mathcal{P}$) admet une solution x^*, que le problème dual ($\mathcal{D}$) admet une solution y^*, et que le saut de dualité

$$\min_{x\in\mathcal{X}}\sup_{y\in\mathcal{Y}} \mathcal{L}(x;y) - \max_{y\in\mathcal{Y}}\inf_{x\in\mathcal{X}} \mathcal{L}(x;y) = J(x^*) - E(y^*)$$

est nul, alors le point (x^*,y^*) est un point-selle de $\mathcal{L}$, c'est-à-dire un point vérifiant

$$\forall\,\big((x^*,y),(x,y^*)\big)\in(\operatorname{dom}\mathcal{L})^2,\qquad \mathcal{L}(x^*,y)\le\mathcal{L}(x^*,y^*)\le\mathcal{L}(x,y^*)$$

En outre, ce point vérifie la condition du premier ordre

$$0\in\partial_x\mathcal{L}(x^*,y^*) \qquad \text{et} \qquad 0\in\partial_y(-\mathcal{L})(x^*,y^*)$$

Les méthodes présentées dans ce module recherche un point-selle d'une fonction de couplage $\mathcal{L}$ en alternant des stratégies de minimisation primale, c'est-à-dire portant sur la variable x, et des stratégies de maximisation duale, c'est-à-dire portant sur la variable y. Plus précisément, elles alternent une itération visant à résoudre le problème

$$\min_{x\in\mathcal{X}} \mathcal{L}(x;y_k)$$

et une itération visant à résoudre le problème

$$\max_{y\in\mathcal{Y}} \mathcal{L}(x_{k+1};y)$$

Ces itérations peuvent être choisies parmi les itérations des méthodes abordées dans ce cours, à savoir, pour la minimisation par exemple : minimisation exacte, minimisation partielle alternée, pas de gradient (explicite, implicite), itération FBS... Pour la maximisation, il suffit de considérer le problème de minimisation équivalent. Le choix de ces itérations dépend des propriétés du problème considéré, et en particulier des propriétés de la fonction de couplage choisie. Dans ce module, on s'intéresse à deux exemples de fonction de couplage, à savoir le lagrangien augmenté et une fonction de couplage formée à l'aide de la conjuguée convexe.

2 Algorithme de Chambolle–Pock CHAMBOLLE-POCK 算法

2.1 Position du problème

On s'intéresse aux problèmes de la forme

原始问题
$$\min_{x\in\mathcal{X}} \left\{ f(A\,x) + g(x) \right\} \tag{$\mathcal{P}_{\text{C-P}}$}$$

où $f : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R}\cup\{+\infty\}$ sont des fonctions convexes, s.c.i. et propres, $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné, de norme $|||A|||$. On suppose également que f et g sont simples. En utilisant la conjuguée convexe de f, on peut écrire que

凸共轭性质
$$\forall\, x\in\mathcal{X}, \qquad f(A\,x) = \sup_{y\in\mathcal{Y}} \left\{ \langle y, A\,x\rangle - f^*(y) \right\}$$

de sorte que le problème $(\mathcal{P}_{\text{C-P}})$ s'exprime comme la recherche d'un point-selle :

寻找鞍点问题
$$\min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \left\{ g(x) + \langle y, A\,x\rangle - f^*(y) \right\}$$

On note $\mathcal{L}$ la fonction de couplage de ce problème.

Intéressons-nous au problème dual associé

对偶问题
$$\max_{y\in\mathcal{Y}} \inf_{x\in\mathcal{X}} \left\{ g(x) + \langle y, A\,x\rangle - f^*(y) \right\}$$

Notons que, pour tout $y\in\mathcal{Y}$, la fonction objectif de ce problème s'écrit

$$\inf_{x\in\mathcal{X}} \left\{ g(x) + \langle y, A\,x\rangle - f^*(y) \right\} = -\sup_{x\in\mathcal{X}} \left\{ -g(x) - \langle A^* y, x\rangle + f^*(y) \right\} = -g^*(-A^* y) - f^*(y)$$

Cette fonction étant concave, la règle de Fermat (pour le problème de minimisation convexe équivalent) assure que toute solution y^* du problème dual vérifie

$$0 \in -A\,\partial(g^*)(-A^* y^*) + \partial(f^*)(y^*)$$

Par ailleurs, écrivons la règle de Fermat pour le problème primal : x^* est solution de ce problème si et seulement si

$$0 \in A^*\,\partial f(A\,x^*) + \partial g(x^*)$$

Si un tel x^* existe, alors, pour tout $p \in \partial f(A\,x^*)$, on a d'après l'inclusion qui précède

$$-A^*\,p \in \partial g(x^*)$$

La règle de bascule garantit alors que

$$x^* \in \partial(g^*)(-A^*\,p) \qquad \text{et} \qquad A\,x^* \in \partial(f^*)(p)$$

Autrement dit, p est une solution du problème dual. Pour démontrer que (x^*, p) est un point-selle de $\mathcal{L}$, il suffit alors de démontrer que l'absence de saut de dualité. Notons que

$$f(A\,x^*) = \langle p, A\,x^*\rangle - f^*(p) \qquad \text{et} \qquad g(x^*) = -\langle A^*\,p, x^*\rangle - g^*(-A^*p)$$

(identité de LEGENDRE–FENCHEL). En sommant ses deux identités, on en déduit que

$$J(x^*) = f(A\,x^*) + g(x^*) = -g^*(-A^*p) - f^*(p) = E(p)$$

ce qui se traduit par un saut de dualité nul. On a donc établi le résultat suivant :

Proposition 1

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire. Soit $(x^*, y^*) \in \mathcal{X} \times \mathcal{Y}$. On a équivalence entre les énoncés suivants :

(i) x^* est un minimiseur de la fonction $J = f \circ A + g$ et y^* un maximiseur de la fonction $E = -g^* \circ (-A^*) - f^*$;

(ii) (x^*, y^*) est un point-selle de la fonction de couplage
鞍点

$$\mathcal{L} : (x,y) \mapsto g(x) + \langle y, A\,x \rangle - f^*(y)$$

2.2 Algorithme

L'algorithme de CHAMBOLLE–POCK pour résoudre le problème $(\mathcal{P}_{\text{C-P}})$ est donné par

$$\begin{cases} (x_0, y_0) \in \mathcal{X} \times \mathcal{Y} \\ \tilde{x}_0 = x_0 \end{cases} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} y_{k+1} = \operatorname{prox}_{\sigma f^*}\left(y_k + \sigma\, A\,\tilde{x}_k\right) \\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau\, A^* y_{k+1}\right) \\ \tilde{x}_{k+1} = x_{k+1} + \theta\,(x_{k+1} - x_k) \end{cases}$$

où $\theta \in [\,0\,;1\,]$ et τ, $\sigma > 0$. Cet algorithme alterne donc

1. une **montée de gradient explicite-implicite** sur la variable duale y ;
2. une **descente de gradient explicite-implicite** sur la variable primale x.

On notera l'ajout d'un pas de sur-relaxation sur la variable primale. Lorsque cette étape n'existe pas ($\theta = 0$), l'algorithme résultant est connu sous le nom de méthode d'ARROW–HURWICZ ou encore PDHG (*Primal-Dual Hybrid Gradient*).

2.3 Propriétés de convergence 收敛性质

Pour étudier les propriétés de l'algorithme de CHAMBOLLE–POCK, on va considérer les itérations générales suivantes :

$$\begin{cases} y^+ = \operatorname{prox}_{\sigma f^*}\left(\bar{y} + \sigma\, A\,\tilde{x}\right) \\ x^+ = \operatorname{prox}_{\tau g}\left(\bar{x} - \tau\, A^*\tilde{y}\right) \end{cases} \tag{$\star$}$$

Dans le cas de l'algorithme de CHAMBOLLE–POCK, chaque $n+1$-ème itération correspond au choix suivant pour les six variables apparaissant dans la formule précédente :

$$\begin{cases} x^+ = x_{k+1} \\ \bar{x} \ \ = x_k \\ \tilde{x} \ \ = x_k + \theta\,(x_k - x_{k-1}) \end{cases} \qquad \text{et} \qquad \begin{cases} y^+ = y_{k+1} \\ \bar{y} \ \ = y_k \\ \tilde{y} \ \ = y_{k+1} \end{cases}$$

Revenons aux itérations ($\star$). La règle de FERMAT écrite pour les opérateurs proximaux implique que

$$\frac{\bar{y} - y^+}{\sigma} + A\,\tilde{x} \in \partial f^*(y^+) \qquad \text{et} \qquad \frac{\bar{x} - x^+}{\tau} - A^*\,\tilde{y} \in \partial g(x^+)$$

On va commencer par établir quelques lemmes préparatoires.

Lemme 1

Soit $f:\mathcal{X}\to\mathbb{R}\cup\{+\infty\}$ et $g:\mathcal{Y}\to\mathbb{R}\cup\{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A:\mathcal{X}\to\mathcal{Y}$ un opérateur linéaire. Soit $(x^+,\bar{x},\tilde{x},y^+,\bar{y},\tilde{y})\in\mathcal{X}^3\times\mathcal{Y}^3$ définis dans $(\star)$. Alors pour tout $(x,y)\in\mathcal{X}\times\mathcal{Y}$, on a

$$\begin{aligned}\frac{1}{2\,\sigma}\,\|y-\bar{y}\|^2+\frac{1}{2\,\tau}\,\|x-\bar{x}\|^2 \geq \mathcal{L}(x^+;y)-\mathcal{L}(x;y^+)&\\ +\frac{1}{2\,\sigma}\,\|y-y^+\|^2+\frac{1}{2\,\tau}\,\|x-x^+\|^2&\\ +\frac{1}{2\,\sigma}\,\|y^+-\bar{y}\|^2+\frac{1}{2\,\tau}\,\|x^+-\bar{x}\|^2&\\ +\langle y^+-\tilde{y},A\,(x-x^+)\rangle+\langle y-y^+,A\,(\tilde{x}-x^+)\rangle&\end{aligned}$$

DÉMONSTRATION : On commence par écrire la définition des sous-gradients pour les fonctions convexes f^* et g :

$$\begin{aligned}&\forall\,y\in\mathcal{Y}, \qquad f^*(y)\geq f^*(y^+)+\frac{1}{\sigma}\,\langle\bar{y}-y^+,y-y^+\rangle+\langle A\,\tilde{x},y-y^+\rangle\\ &\forall\,x\in\mathcal{X}, \qquad g(x)\geq g(x^+)+\frac{1}{\tau}\,\langle\bar{x}-x^+,x-x^+\rangle-\langle\tilde{y},A\,(x-x^+)\rangle\end{aligned}$$

En ajoutant respectivement les quantités $\|y-\bar{y}\|^2/(2\,\sigma)$ et $\|x-\bar{x}\|^2/(2\,\tau)$ dans les deux inégalités précédentes, puis en développant les carrés dans les membres de droite, on obtient après simplification que, pour tout $x\in\mathcal{X}$,

$$f^*(y)+\frac{1}{2\,\sigma}\,\|y-\bar{y}\|^2\geq f^*(y^+)+\langle A\,\tilde{x},y-y^+\rangle+\frac{1}{2\,\sigma}\,\|y-y^+\|^2+\frac{1}{2\,\sigma}\,\|y^+-\bar{y}\|^2$$

et que, pour tout $y\in\mathcal{Y}$,

$$g(x)+\frac{1}{2\,\tau}\,\|x-\bar{x}\|^2\geq g(x^+)-\langle\tilde{y},A\,(x-x^+)\rangle+\frac{1}{2\,\tau}\,\|x-x^+\|^2+\frac{1}{2\,\tau}\,\|x^+-\bar{x}\|^2$$

Notons que, pour tout $(x,y)\in\mathcal{X}\times\mathcal{Y}$, on a l'identité

$$\langle y^+,A\,x\rangle-\langle y,A\,x^+\rangle=\langle y^+,A\,(x-x^+)\rangle-\langle y-y^+,A\,x^+\rangle$$

Ainsi, en sommant les deux inégalités précédentes, on obtient après réarrangement des termes l'inégalité annoncée. ■

Remplaçons les variables génériques $(x^+,\bar{x},\tilde{x},y^+,\bar{y},\tilde{y})$ par leur expression dans l'algorithme de CHAMBOLLE–POCK ; on a en particulier

$$\begin{cases}x^+-\bar{x}=x_{k+1}-x_k\\ \tilde{x}-x^+=x_k-x_{k+1}+\theta\,(x_k-x_{k-1})\end{cases}\qquad\text{et}\qquad\begin{cases}y^+-\bar{y}=y_{k+1}-y_k\\ \tilde{y}-y^+=0\end{cases}$$

de sorte que l'inégalité démontrée dans le lemme 2 devient

$$\begin{aligned}\frac{1}{2\,\sigma}\,\|y-y_k\|^2+\frac{1}{2\,\tau}\,\|x-x_k\|^2\geq\mathcal{L}(x_{k+1};y)-\mathcal{L}(x;y_{k+1})&\\ +\frac{1}{2\,\sigma}\,\|y-y_{k+1}\|^2+\frac{1}{2\,\tau}\,\|x-x_{k+1}\|^2&\\ +\frac{1}{2\,\sigma}\,\|y_{k+1}-y_k\|^2+\frac{1}{2\,\tau}\,\|x_{k+1}-x_k\|^2&\\ +\langle y-y_{k+1},A\,(x_k-x_{k+1}+\theta\,(x_k-x_{k-1}))\rangle&\end{aligned}$$

On va commencer par minorer le produit scalaire qui apparaît dans l'inégalité précédente.

Lemme 2

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné, de norme $|||A|||$. Considérons la suite $((x_k, \tilde{x}_k, y_k))_{k\in\mathbb{N}} \in (\mathcal{X}^2 \times \mathcal{Y})^{\mathbb{N}}$ générée par l'algorithme

$$\begin{cases} (x_0, y_0) \in \mathcal{X} \times \mathcal{Y} \\ \tilde{x}_0 = x_0 \end{cases} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} y_{k+1} = \operatorname{prox}_{\sigma f^*}\left(y_k + \sigma\, A\, \tilde{x}_k\right) \\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau\, A^* y_{k+1}\right) \\ \tilde{x}_{k+1} = x_{k+1} + \theta\,(x_{k+1} - x_k) \end{cases}$$

où $\theta \in [\,0\,;1\,]$ et $\tau,\, \sigma > 0$. Alors pour tout $n \in \mathbb{N}$, on a pour tout $\omega \in [\,0\,;\theta\,]$

$$\begin{aligned} &\langle y - y_{k+1}, A\,(x_k - x_{k+1} + \theta\,(x_k - x_{k-1}))\rangle \\ &\qquad \geq \langle y - y_{k+1}, A\,(x_k - x_{k+1})\rangle - \omega\,\langle y - y_k, A\,(x_{k-1} - x_k)\rangle \\ &\qquad\quad - \frac{1}{2\,\sigma}\,\|y_k - y_{k+1}\|^2 - \frac{\omega\,\theta\,|||A|||^2\,\tau\,\sigma}{2\,\tau}\,\|x_{k-1} - x_k\|^2 \\ &\qquad\quad - \frac{\theta - \omega}{\omega}\,\frac{1}{2\,\sigma}\,\|y - y_{k+1}\|^2 \end{aligned}$$

Démonstration : On commence par écrire que, pour tout $\omega \in [\,0\,;\theta\,]$,

$$\begin{aligned} \theta\,\langle y - y_{k+1}, A\,(x_{k-1} - x_k)\rangle = {} & \omega\,\langle y - y_k, A\,(x_{k-1} - x_k)\rangle \\ & + \omega\,\langle y_k - y_{k+1}, A\,(x_{k-1} - x_k)\rangle \\ & + (\theta - \omega)\,\langle y - y_{k+1}, A\,(x_{k-1} - x_k)\rangle \end{aligned}$$

Ainsi, en utilisant l'inégalité de Cauchy–Schwarz et l'inégalité $2\,a\,b \leq a^2/\beta + \beta\,b^2$ valable pour tout $\beta > 0$, on a d'une autre part

$$\begin{aligned} \langle y_k - y_{k+1}, A\,(x_{k-1} - x_k)\rangle &\leq |||A||| \cdot \|y_k - y_{k+1}\| \cdot \|x_{k-1} - x_k\| \\ &\leq |||A|||\left(\frac{\sigma/\beta}{2\,\sigma}\,\|y_k - y_{k+1}\|^2 + \frac{\beta\,\tau}{2\,\tau}\,\|x_{k-1} - x_k\|^2\right) \end{aligned}$$

et, d'autre part, en procédant de manière similaire,

$$\langle y - y_{k+1}, A\,(x_{k-1} - x_k)\rangle \leq |||A|||\left(\frac{\sigma/\beta}{2\,\sigma}\,\|y - y_{k+1}\|^2 + \frac{\beta\,\tau}{2\,\tau}\,\|x_{k-1} - x_k\|^2\right)$$

Il s'ensuit que

$$\begin{aligned} &\omega\,\langle y_k - y_{k+1}, A\,(x_{k-1} - x_k)\rangle + (\theta - \omega)\,\langle y - y_{k+1}, A\,(x_{k-1} - x_k)\rangle \\ &\qquad \leq \frac{\omega\,|||A|||\,\sigma/\beta}{2\,\sigma}\,\|y_k - y_{k+1}\|^2 + \frac{\theta\,|||A|||\,\beta\,\tau}{2\,\tau}\,\|x_{k-1} - x_k\|^2 \\ &\qquad\quad + \frac{(\theta - \omega)\,|||A|||\,\sigma/\beta}{2\,\sigma}\,\|y - y_{k+1}\|^2 \end{aligned}$$

On choisit $\beta = \omega\,|||A|||\,\sigma$. Après simplification, on obtient que

$$\begin{aligned} &\omega\,\langle y_k - y_{k+1}, A\,(x_{k-1} - x_k)\rangle + (\theta - \omega)\,\langle y - y_{k+1}, A\,(x_{k-1} - x_k)\rangle \\ &\qquad \leq \frac{1}{2\,\sigma}\,\|y_k - y_{k+1}\|^2 + \frac{\omega\,\theta\,|||A|||^2\,\tau\,\sigma}{2\,\tau}\,\|x_{k-1} - x_k\|^2 \\ &\qquad\quad + \frac{\theta - \omega}{\omega}\,\frac{1}{2\,\sigma}\,\|y - y_{k+1}\|^2 \end{aligned}$$

ce qui est bien le résultat annoncé. ■

Posons $$\forall\, k \in \mathbb{N}, \qquad \Delta_k = \frac{1}{2\,\sigma}\,\|y - y_k\|^2 + \frac{1}{2\,\tau}\,\|x - x_k\|^2$$

Lemme 3 (Cas $\theta = 1$)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné, de norme $|||A|||$. Considérons la suite $((x_k, \tilde{x}_k, y_k))_{k\in\mathbb{N}} \in (\mathcal{X}^2 \times \mathcal{Y})^{\mathbb{N}}$ générée par l'algorithme

$$\begin{cases} (x_0, y_0) \in \mathcal{X} \times \mathcal{Y} \\ \tilde{x}_0 = x_0 = x_{-1} \end{cases} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} y_{k+1} = \operatorname{prox}_{\sigma f^*}\left(y_k + \sigma\, A\, \tilde{x}_k\right) \\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau\, A^* y_{k+1}\right) \\ \tilde{x}_{k+1} = 2\, x_{k+1} - x_k \end{cases}$$

où τ, $\sigma > 0$. Alors on a pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$

$$\begin{aligned} \frac{1}{2\,\sigma}\,\|y - y_0\|^2 + \frac{1}{2\,\tau}\,\|x - x_0\|^2 \geq{} & \sum_{k=0}^{K-1} \mathcal{L}(x_{k+1}; y) - \sum_{k=0}^{K-1} \mathcal{L}(x; y_{k+1}) \\ & + \frac{1 - \tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y - y_K\|^2 + \frac{1}{2\,\tau}\,\|x - x_K\|^2 \\ & + (1 - |||A|||\,\sqrt{\tau\,\sigma}) \sum_{k=0}^{K-1} \left(\frac{1}{2\,\sigma}\,\|y_{k+1} - y_k\|^2 + \frac{1}{2\,\tau}\,\|x_{k+1} - x_k\|^2 \right) \end{aligned}$$

Démonstration : On combine les résultats des lemmes 3 et 4 dans le cas où on choisit $\theta = \omega = 1$:

$$\begin{aligned} \Delta_k \geq{} & \mathcal{L}(x_{k+1}; y) - \mathcal{L}(x; y_{k+1}) + \Delta_{k+1} + \frac{1}{2\,\tau}\,\|x_{k+1} - x_k\|^2 - \frac{1}{2\,\tau}\,\|x_k - x_{k-1}\|^2 \\ & + \frac{1 - |||A|||\,\sqrt{\tau\,\sigma}}{2\,\sigma}\,\|y_{k+1} - y_k\|^2 + \frac{1 - |||A|||\,\sqrt{\tau\,\sigma}}{2\,\tau}\,\|x_k - x_{k-1}\|^2 \\ & + \langle y - y_{k+1}, A\,(x_k - x_{k+1})\rangle - \langle y - y_k, A\,(x_{k-1} - x_k)\rangle \end{aligned}$$

On somme pour k entre 0 et $K - 1$, puis on divise par K ; puisque $x_{-1} = x_0$, on a

$$\begin{aligned} \Delta_0 \geq{} & \sum_{k=0}^{K-1} \mathcal{L}(x_{k+1}; y) - \sum_{k=0}^{K-1} \mathcal{L}(x; y_{k+1}) + \Delta_K + \frac{1}{2\,\tau}\,\|x_K - x_{K-1}\|^2 \\ & + (1 - |||A|||\,\sqrt{\tau\,\sigma}) \sum_{k=0}^{K-1} \left(\frac{1}{2\,\sigma}\,\|y_{k+1} - y_k\|^2 + \frac{1}{2\,\tau}\,\|x_{k+1} - x_k\|^2 \right) \\ & + \langle y - y_K, A\,(x_{K-1} - x_K)\rangle \end{aligned}$$

tandis que, grâce à l'inégalité de Cauchy–Schwarz, on a à nouveau (en choisissant cette fois $\beta = 1/(\tau\,|||A|||)$)

$$\langle y_K - y, A\,(x_K - x_{K-1})\rangle \leq \frac{\tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y_K - y\|^2 + \frac{1}{2\,\tau}\,\|x_K - x_{K-1}\|^2$$

On combine les deux dernières inégalités pour obtenir le résultat annoncé. ■

Puisque les fonctions partielles $x \mapsto \mathcal{L}(x; y)$ et $y \mapsto -\mathcal{L}(x; y)$ sont convexes, on a pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$

$$\frac{1}{K} \sum_{k=0}^{K-1} \mathcal{L}(x_{k+1}; y) \geq \mathcal{L}\left(\frac{1}{K} \sum_{k=0}^{K-1} x_{k+1}; y \right) \quad \text{et} \quad \frac{1}{K} \sum_{k=0}^{K-1} \mathcal{L}(x; y_{k+1}) \leq \mathcal{L}\left(x; \frac{1}{K} \sum_{k=0}^{K-1} y_{k+1} \right)$$

Considérons les moyennes de CESÀRO de cette suite, définies par

$$\forall\, K \in \mathbb{N}^*, \qquad X_K = \frac{1}{K} \sum_{k=0}^{K-1} x_{k+1} \qquad \text{et} \qquad Y_K = \frac{1}{K} \sum_{k=0}^{K-1} y_{k+1}$$

de sorte que le lemme 3 s'écrit

$$\begin{aligned} \frac{1}{2\,\sigma}\,\|y-y_0\|^2 + \frac{1}{2\,\tau}\,\|x-x_0\|^2 &\geq K\,\mathcal{L}(X_K;y) - K\,\mathcal{L}(x;Y_K) \\ &\quad + \frac{1-\tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y-y_K\|^2 + \frac{1}{2\,\tau}\,\|x-x_K\|^2 \\ + (1-|||A|||\,\sqrt{\tau\,\sigma}) &\sum_{k=0}^{K-1} \left(\frac{1}{2\,\sigma}\,\|y_{k+1}-y_k\|^2 + \frac{1}{2\,\tau}\,\|x_{k+1}-x_k\|^2 \right) \end{aligned}$$

Proposition 2 (Cas $\theta = 1$)

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné, de norme $|||A|||$. Considérons la suite $((x_k, \tilde{x}_k, y_k))_{k \in \mathbb{N}} \in (\mathcal{X}^2 \times \mathcal{Y})^{\mathbb{N}}$ générée par l'algorithme

$$\begin{cases} (x_0, y_0) \in \mathcal{X} \times \mathcal{Y} \\ \tilde{x}_0 = x_0 = x_{-1} \end{cases} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} y_{k+1} = \operatorname{prox}_{\sigma\, f^*}\left(y_k + \sigma\, A\, \tilde{x}_k\right) \\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau\, A^* y_{k+1}\right) \\ \tilde{x}_{k+1} = 2\,x_{k+1} - x_k \end{cases}$$

où τ, $\sigma > 0$. Si $\tau\,\sigma < 1/|||A|||^2$, alors on a

$$\lim_{k \to +\infty} \|x_{k+1} - x_k\| = \lim_{k \to +\infty} \|y_{k+1} - y_k\| = 0$$

DÉMONSTRATION : Il suffit d'appliquer l'inégalité du lemme 3 à un point-selle $(\hat{x}, \hat{y})$ de $\mathcal{L}$; puisque

$$\mathcal{L}(X_K; \hat{y}) - \mathcal{L}(\hat{x}; Y_K) \geq 0$$

et que $1 - |||A|||\,\sqrt{\tau\,\sigma} > 0$, on en déduit que les séries de terme général $\|y_{k+1} - y_k\|^2$ et $\|x_{k+1} - x_k\|^2$ sont convergentes ; on en déduit que les deux suites associées au terme général tendent vers 0. ■

Proposition 3 (Cas $\theta = 1$) θ = 1 时，生成序列的收敛性

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Y} \to \mathbb{R} \cup \{+\infty\}$ des fonctions **convexes**, s.c.i. et propres, et $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné, de norme $|||A|||$. Considérons la suite $((x_k, \tilde{x}_k, y_k))_{k \in \mathbb{N}} \in (\mathcal{X}^2 \times \mathcal{Y})^{\mathbb{N}}$ générée par l'algorithme

$$\begin{cases} (x_0, y_0) \in \mathcal{X} \times \mathcal{Y} \\ \tilde{x}_0 = x_0 = x_{-1} \end{cases} \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} y_{k+1} = \operatorname{prox}_{\sigma\, f^*}\left(y_k + \sigma\, A\, \tilde{x}_k\right) \\ x_{k+1} = \operatorname{prox}_{\tau g}\left(x_k - \tau\, A^* y_{k+1}\right) \\ \tilde{x}_{k+1} = 2\,x_{k+1} - x_k \end{cases}$$

où τ, $\sigma > 0$. Si $\tau\,\sigma \leq 1/|||A|||^2$, alors la suite $(x_k)_{k \in \mathbb{N}}$ est bornée. De plus, si $\tau\,\sigma < 1/|||A|||^2$, alors la suite $((x_k, y_k))_{k \in \mathbb{N}}$ converge vers un point-selle de $\mathcal{L}$. En particulier, la suite $(x_k)_{k \in \mathbb{N}}$ converge vers une solution du problème

$$\min_{x \in \mathcal{X}} \left\{ f(A\,x) + g(x) \right\}$$

Démonstration : Décomposons la preuve en plusieurs étapes.

- **La suite $(x_k)_{k\in\mathbb{N}}$ est bornée.** On reprend l'inégalité du lemme 3; on a pour tout $K \in \mathbb{N}^*$

$$\begin{aligned}\Delta_0 &\geq K\,\left(\mathcal{L}(X_K;\hat{y}) - \mathcal{L}(\hat{x};Y_K)\right) + \Delta_K - \frac{\tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y_K - y\|^2 \\ &\geq \frac{1-\tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|\hat{y} - y_K\|^2 + \frac{1}{2\,\tau}\,\|\hat{x} - x_K\|^2 \geq 0\end{aligned}$$

car $|||A|||\,\sqrt{\tau\,\sigma} \leq 1$ et que $(\hat{x},\hat{y})$ est un point-selle de $\mathcal{L}$. On en déduit que la suite $(x_k)_{k\in\mathbb{N}}$ est bornée. De plus, si $|||A|||\sqrt{\tau\sigma} < 1$, alors la suite $(y_k)_{k\in\mathbb{N}}$ est également bornée.

- **Toute valeur d'adhérence de $((X_K,Y_K))_{K\in\mathbb{N}^*}$ est un point-selle de $\mathcal{L}$.** On suppose que $|||A|||\sqrt{\tau\sigma} < 1$. Il est aisé de vérifier que les suites $(X_K)_{K\in\mathbb{N}^*}$ et $(Y_K)_{K\in\mathbb{N}^*}$ sont bornées, et que la suite $((X_K,Y_K))_{K\in\mathbb{N}^*}$ admet donc une sous-suite $((X_{K_j},Y_{K_j}))_{j\in\mathbb{N}}$ qui converge, vers un point noté (X^*,Y^*). Par ailleurs, on a démontré dans le lemme 3 que, pour tout $K\in\mathbb{N}^*$,

$$\frac{\Delta_0}{K} \geq \mathcal{L}(X_K;y) - \mathcal{L}(x;Y_K)$$

En appliquant cette relation à la sous-suite $((X_{K_j},Y_{K_j}))_{j\in\mathbb{N}}$, puis en faisant tendre j vers $+\infty$, il s'ensuit par encadrement que, puisque g et f^* sont s.c.i.,

$$0 \geq \lim_{j\to+\infty}\mathcal{L}(X_{K_j};y) - \lim_{j\to+\infty}\mathcal{L}(x;Y_{K_j}) \geq \mathcal{L}(X^*;y) - \mathcal{L}(x;Y^*)$$

Cette relation étant valable pour tout (X^*,y) et (x,Y^*), on en déduit que le point (X^*,Y^*) est un point-selle de $\mathcal{L}$.

- **Toute valeur d'adhérence de $((x_k,y_k))_{k\in\mathbb{N}}$ est un point-selle de $\mathcal{L}$.** Soit $((x_{k_j},y_{k_j}))_{j\in\mathbb{N}}$ une sous-suite convergente, de limite (x^*,y^*). On a établi avant le lemme 3 que, pour tout $j\in\mathbb{N}$,

$$\frac{y_{k_j} - y_{k_j+1}}{\sigma} + A\,(x_{k_j-1} - x_{k_j}) \in -A\,x_{k_j} + \partial(f^*)(y_{k_j+1}) = \partial_y(-\mathcal{L})(x_{k_j};y_{k_j+1})$$

et que
$$\frac{x_{k_j-1} - x_{k_j}}{\tau} \in A^*\,y_{k_j+1} + \partial g(x_{k_j+1}) = \partial_x\mathcal{L}(x_{k_j+1};y_{k_j+1})$$

puisque les suites $(\|x_{k_j-1} - x_{k_j}\|)_{j\in\mathbb{N}}$ et que $(\|y_{k_j} - y_{k_j+1}\|)_{j\in\mathbb{N}}$ tendent vers 0, les suites $((x_{k_j+1},y_{k_j+1}))_{j\in\mathbb{N}}$ tendent vers (x^*,y^*). Ainsi, le passage à la limite assure que, par fermeture des sous-différentiels de f^* et de g,

$$0 \in \partial_x\mathcal{L}(x^*;y^*) \qquad \text{et} \qquad 0 \in \partial_y(-\mathcal{L})(x^*;y^*)$$

La proposition 4 du module **A6 : Dualité min-max**permet de démontrer que (x^*,y^*) est un point-selle de $\mathcal{L}$.

- **La suite $((x_k,y_k))_{k\in\mathbb{N}}$ converge vers (x^*,y^*).** Soit $\varepsilon > 0$. Il existe $j_0\in\mathbb{N}$ tel que, pour tout $j \geq j_0$,

$$\frac{1}{2\,\sigma}\,\|y^* - y_{k_j}\|^2 + \frac{1}{2\,\tau}\,\|x^* - x_{k_j}\|^2 \leq \varepsilon$$

Soit $K \geq k_{j_0}$. En appliquant le lemme 3 aux indices K et k_{j_0}, on obtient après soustraction que

$$\begin{aligned}\frac{1}{2\,\sigma}\,\|y^* - y_{k_j}\|^2 + \frac{1}{2\,\tau}\,\|x^* - x_{k_j}\|^2 &\geq \sum_{k=k_j}^{K-1}\mathcal{L}(x_{k+1};y^*) - \sum_{k=k_j}^{K-1}\mathcal{L}(x^*;y_{k+1}) \\ &\quad + \frac{1-\tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y^* - y_K\|^2 + \frac{1}{2\,\tau}\,\|x^* - x_K\|^2 \\ + (1 - |||A|||\,\sqrt{\tau\,\sigma})&\sum_{k=k_j}^{K-1}\left(\frac{1}{2\,\sigma}\,\|y_{k+1} - y_k\|^2 + \frac{1}{2\,\tau}\,\|x_{k+1} - x_k\|^2\right)\end{aligned}$$

Puisque (x^*, y^*) est un point-selle de $\mathcal{L}$, on a pour tout $k \in \mathbb{N}$

$$\mathcal{L}(x_{k+1}; y^*) - \mathcal{L}(x^*; y_{k+1}) \geq 0$$

Il s'ensuit que

$$\varepsilon \geq \frac{1 - \tau\,\sigma\,|||A|||^2}{2\,\sigma}\,\|y^* - y_K\|^2 + \frac{1}{2\,\tau}\,\|x^* - x_K\|^2$$

Autrement dit, les suites $(x_K)_{K\in\mathbb{N}^*}$ et $(y_K)_{K\in\mathbb{N}^*}$ tendent vers x^* et y^* respectivement. ■

> *Cas régulier.* Lorsque la fonction g est fortement convexe de module α et la fonction f $L_{\nabla f}$-régulière, il est aisé de vérifier que $f \circ A$ est $|||A|||^2\,L_{\nabla f}$-régulière. Dans ce cas, il est possible de choisir un $\theta < 1$ dans l'algorithme de Chambolle–Pock, et démontrer une convergence linéaire, dont le taux de convergence théorique dépend des paramètres du problème.

2.4 Exemple d'applications

Un des modèles les plus simples et les plus connus de débruitage d'images est le modèle de débruitage pur de Rudin–Osher–Fatemi (ROF). 图像去噪 Il consiste, dans le cas d'une image en niveaux de gris, à résoudre le problème de minimisation suivant :

$$\min_{u\in\mathbb{R}^{m\times n}} \frac{1}{2}\,\|u - v\|^2 + \lambda\,\mathrm{TV}(u)$$

où les données v sont une version dégradée d'un signal u que l'on cherche à reconstruire avec

$$v = u + n$$

Dans ce modèle, n le bruit additif blanc gaussien. Le premier terme est appelé *terme d'attache aux données* ou *terme de fidélité* ; il sert à imposer à la solution u^* à rester proche des données bruitées v. Le second terme est le terme de régularisation : ici, la régularisation utilisée est la régularisation TV (pour Variation Totale), définie comme la norme du gradient. Une manière d'approcher de façon discrète cette norme est d'introduire une version discrète du gradient :

离散形式的梯度

$$\nabla^h u = (\delta_x^h u, \delta_y^h u)$$

avec $\forall\,(i,j) \in [\![\,1\,;\,m\,]\!] \times [\![\,1\,;\,n\,]\!], \qquad (\delta_x^h u)_{(i,j)} = \begin{cases} u_{(i+1,j)} - u_{(i,j)} & \text{si } i \neq m \\ 0 & \text{sinon} \end{cases}$

et $\forall\,(i,j) \in [\![\,1\,;\,m\,]\!] \times [\![\,1\,;\,n\,]\!], \qquad (\delta_y^h u)_{(i,j)} = \begin{cases} u_{(i,j+1)} - u_{(i,j)} & \text{si } j \neq n \\ 0 & \text{sinon} \end{cases}$

Dans ce cas, on définit $\qquad \mathrm{TV}(u) = \|\nabla^h u\|_2$

Enfin, le réel positif λ est un paramètre qui permet de moduler l'importance relative des deux termes ; plus λ est grand, plus le terme de régularisation est prépondérant, et de manière équivalente moins le terme d'attache aux données est fort ; en particulier, cela signifie que la solution u^* sera régulier, au détriment de sa proximité avec la donnée bruitée v.

Notons J la fonction objectif de ce problème. Cette fonction est convexe. Le terme de fidélité g est simple. En effet,

$$u^+ = \mathrm{prox}_{\tau f}(u^0) = \operatorname*{argmin}_{u\in\mathbb{R}^{m\times n}} \left\{ \underbrace{\frac{1}{2}\,\|u - g\|_2^2}_{g(u)} + \underbrace{\frac{1}{2\,\tau}\,\|u - u^0\|_2^2}_{\frac{1}{2}\|\nabla^h u\|_2^2} \right\}$$

$f(\nabla^h u)$

La règle de FERMAT s'écrit alors

$$u^+ - g + \frac{1}{\tau}\,(u^+ - u^0) = 0 \qquad \text{soit} \qquad u^+ = \frac{u^0 + \tau\, g}{1+\tau}$$

$\text{prox}_{\sigma f^*} = \text{proj}_{\overline{B(0,1)}}$

Intéressons-nous à présent au terme de régularité, de la forme $f(\nabla^h u)$ avec f la norme euclidienne multipliée par λ. Ce terme n'est pas simple, mais f l'est. En effet, en utilisant l'identité de MOREAU généralisée, on prouve que $\text{prox}_{\sigma f^*}$ est la projection orthogonale sur la boule unité. Les itérations de l'algorithme de CHAMBOLLE–POCK s'écrivent alors

$$\forall\, k \in \mathbb{N}, \qquad \begin{cases} v_{k+1} = \text{proj}_{\mathcal{B}(0,1)}\left(v_k + \sigma\, \nabla^h \bar{u}_k\right) \\ u_{k+1} = \dfrac{u_k - \tau\,(\nabla^h)^* v_{k+1} + \tau\, g}{1+\tau} \\ \bar{u}_{k+1} = u_{k+1} + \theta\,(u_{k+1} - u_k) \end{cases}$$

3 Méthode des directions alternées 交替方向法

3.1 Position du problème

On s'intéresse aux problèmes de la forme

$$\min_{\substack{(x,z)\in\mathcal{X}\times\mathcal{Z} \\ Ax+Bz=c}} \Big\{ f(x) + g(z) \Big\} \tag{$\mathcal{P}_{\text{ADMM}}$}$$

où $f : \mathcal{X} \to \mathbb{R}\cup\{+\infty\}$ et $g : \mathcal{Z} \to \mathbb{R}\cup\{+\infty\}$ sont des fonctions convexes, s.c.i. et propres, $A : \mathcal{X} \to \mathcal{Y}$ et $B : \mathcal{Z} \to \mathcal{Y}$ deux opérateurs linéaires bornés et $c \in \mathcal{Y}$ un vecteur.

Le lagrangien augmenté associé est donné pour $\tau > 0$ par

$$\forall\,(x,z,\lambda) \in \mathcal{X}\times\mathcal{Z}\times\mathcal{Y}, \quad \mathcal{L}_\tau(x,z;\lambda) = f(x)+g(z)+\langle \lambda, A\,x+B\,z-c\rangle+\frac{1}{2\,\tau}\,\|A\,x+B\,z-c\|^2$$

Rappelons que lagrangien et lagrangien augmenté ont mêmes points-selles. Par ailleurs, d'après le théorème de KARUSH–KUHN–TUCKER (**A6 : Dualité min-max**), si le problème $(\mathcal{P}_{\text{ADMM}})$ satisfait une condition de qualification (ce qui est bien le cas car le problème est convexe et que les contraintes sont affines, dont la condition de SLATER s'applique), le lagrangien admet un point-selle si le problème primal $(\mathcal{P}_{\text{ADMM}})$ admet une solution. On en déduit que le lagrangien augmenté admet un point-selle et, par convexité, que tout point-selle est associé à un minimiseur du problème primal. Le problème $(\mathcal{P}_{\text{ADMM}})$ s'écrit donc comme le problème de recherche d'un point-selle :

$$\min_{(x,z)\in\mathcal{X}}\ \sup_{\lambda\in\mathcal{Y}} \left\{ f(x) + g(z) + \langle \lambda, A\,x + B\,z - c\rangle + \frac{1}{2\,\tau}\,\|A\,x+B\,z-c\|^2 \right\}$$

On suppose que, pour tout λ fixé, un minimiseur des fonctions partielles $x' \mapsto \mathcal{L}(x',z;\lambda)$ et $z' \mapsto \mathcal{L}(x,z';\lambda)$ pour tout $(x,z) \in \mathcal{X}\times\mathcal{Z}$ existe et que l'on sait le calculer.

3.2 Algorithme ADMM 交替方向乘子 算法

Initialement proposée par GLOWINSKI et MAROCCO en 1975 et GABAY et MERCIER en 1976, la méthode des directions alternées, ou encore *Alternating Direction Method of Multipliers* (ADMM) en anglais, propose de chercher un point-selle du lagrangien augmenté avec les itérations suivantes :

$$(x_0, z_0, \lambda_0) \in \mathcal{X}\times\mathcal{Z}\times\mathcal{Y} \qquad \text{et} \qquad \forall\, k\in\mathbb{N}, \quad \begin{cases} x_{k+1} \in \ \underset{x\in\mathcal{X}}{\text{argmin}}\ \mathcal{L}_\tau(x, z_k; \lambda_k) \\ z_{k+1} \in \ \underset{z\in\mathcal{Z}}{\text{argmin}}\ \mathcal{L}_\tau(x_{k+1}, z; \lambda_k) \\ \lambda_{k+1} = \ \lambda_k + \dfrac{1}{\tau}\,\dfrac{\partial\mathcal{L}_\tau}{\partial\lambda}(x_{k+1}, z_{k+1}; \lambda_k) \end{cases}$$

Ainsi, la méthode ADMM alterne

1. une **minimisation alternée** sur les variables primales x et z ;
2. une **montée de gradient explicite** sur les multiplicateurs de LAGRANGE.

Les deux mises-à-jour primales s'écrivent de manière plus explicite

$$\begin{cases} x_{k+1} \in & \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ f(x) + \langle \lambda_k, A\,x \rangle + \dfrac{1}{2\,\tau} \, \|A\,x + B\,z_k - c\|^2 \right\} \\ z_{k+1} \in & \underset{z \in \mathcal{Z}}{\operatorname{argmin}} \left\{ g(z) + \langle \lambda_k, B\,z \rangle + \dfrac{1}{2\,\tau} \, \|A\,x_{k+1} + B\,z - c\|^2 \right\} \end{cases}$$

tandis que la mise-à-jour des multiplicateurs s'écrit

$$\lambda_{k+1} = \lambda_k + \frac{1}{\tau} \, (A\,x_{k+1} + B\,z_{k+1} - c)$$

Notons que cet algorithme ne génère *a priori* pas des points admissibles, c'est-à-dire qu'en général, on a

$$A\,x_{k+1} + B\,z_{k+1} \neq c$$

On verra que cette identité est réalisée à convergence (lorsque les conditions sont réunies pour la convergence). Or, numériquement, on n'atteint en général pas ce stade ; aussi, l'ADMM permet plutôt de déterminer une approximation de la valeur optimale atteinte dans le problème $(\mathcal{P}_{\text{ADMM}})$, plutôt que pour obtenir des estimations des solutions (en effet, ces estimations n'étant pas admissibles, elles peuvent, dans certaines applications, ne pas être pertinentes).

> Pour résoudre le problème $(\mathcal{P}_{\text{ADMM}})$, on pourrait considérer une variante de l'algorithme ADMM où les itérations primales sont remplacées par
>
> $$(x_{k+1}, z_{k+1}) \in \underset{(x,z) \in \mathcal{X} \times \mathcal{Z}}{\operatorname{argmin}} \; \mathcal{L}_\tau(x, z; \lambda_k)$$
>
> Cette méthode est appelée *méthode du lagrangien augmenté*. Elle est en général plus difficile à mettre en œuvre que l'ADMM. Cette dernière peut en effet être vue comme une variante de la méthode du lagrangien augmenté, dans laquelle on a remplacé la minimisation primale (donnée au-dessus) par un pas de minimisation alternée en x et en z (soit une itération de l'algorithme BCD – voir **B6 : Éclatement de variables**) appliqué à ce problème de minimisation.

3.3 Propriétés de convergence

On va maintenant démontrer qu'il existe un lien entre l'ADMM et un autre algorithme, celui de CHAMBOLLE–POCK. Rappelons l'expression des itérations de l'ADMM :

$$\begin{cases} x_{k+1} \in & \underset{x \in \mathcal{X}}{\operatorname{argmin}} \left\{ f(x) + \langle \lambda_k, A\,x \rangle + \dfrac{1}{2\,\tau} \, \|A\,x + B\,z_k - c\|^2 \right\} \\ z_{k+1} \in & \underset{z \in \mathcal{Z}}{\operatorname{argmin}} \left\{ g(z) + \langle \lambda_k, B\,z \rangle + \dfrac{1}{2\,\tau} \, \|A\,x_{k+1} + B\,z - c\|^2 \right\} \\ \lambda_{k+1} = & \lambda_k + \dfrac{1}{\tau} \, (A\,x_{k+1} + B\,z_{k+1} - c) \end{cases}$$

Considérons les deux fonctions

$$F : \left\{ \begin{array}{ccl} \mathcal{Y} & \to & \mathbb{R} \cup \{+\infty\} \\ \xi & \mapsto & \inf \left\{ f(x) \mid x \in \mathcal{X} \text{ tel que } A\,x - c = \xi \right\} \end{array} \right.$$

et $$G : \left\{ \begin{array}{ccl} \mathcal{Y} & \to & \mathbb{R} \cup \{+\infty\} \\ \nu & \mapsto & \inf \Big\{ g(z) \mid z \in \mathcal{Z} \text{ tel que } -B\,z = \nu \Big\} \end{array} \right.$$

Notons que ces deux fonctions prennent une valeur finie lorsque $\xi + c$ (resp. ν) appartient à l'image de A (resp. $-B$), et valent $+\infty$ sinon. On peut démontrer que ces deux fonctions sont convexes si f et g le sont. Intéressons-nous à la définition de x_{k+1}. Par optimalité, on a pour tout $x \in \mathcal{X}$

$$f(x_{k+1}) + \langle \lambda_k, A\,x_{k+1} \rangle + \frac{1}{2\,\tau}\,\|A\,x_{k+1} + B\,z_k - c\|^2 \leq f(x) + \langle \lambda_k, A\,x \rangle + \frac{1}{2\,\tau}\,\|A\,x + B\,z_k - c\|^2$$

Ainsi, si on pose $\xi_{k+1} = A\,x_{k+1} - c$ et $\xi = A\,x - c$, alors cette relation se lit

$$F(\xi_{k+1}) + \langle \lambda_k, \xi_{k+1} + c \rangle + \frac{1}{2\,\tau}\,\|\xi_{k+1} + B\,z_k\|^2 \leq F(\xi) + \langle \lambda_k, \xi + c \rangle + \frac{1}{2\,\tau}\,\|\xi + B\,z_k\|^2$$

où $\xi + c \in \mathrm{Im} A$. Puisque $F(\xi) = +\infty$ si $\xi + c \notin \mathrm{Im} A$, on en déduit que cette relation est vraie pour tout $\xi \in \mathcal{Y}$. Autrement dit,

$$F(\xi_{k+1}) + \langle \lambda_k, \xi_{k+1} + c \rangle + \frac{1}{2\,\tau}\,\|\xi_{k+1} + B\,z_k\|^2 = \min_{\xi \in \mathcal{Y}} \left\{ F(\xi) + \langle \lambda_k, \xi + c \rangle + \frac{1}{2\,\tau}\,\|\xi + B\,z_k\|^2 \right\}$$

En particulier, on a

$$\xi_{k+1} \in \operatorname*{argmin}_{\xi \in \mathcal{Y}} \left\{ F(\xi) + \langle \lambda_k, \xi + c \rangle + \frac{1}{2\,\tau}\,\|\xi + B\,z_k\|^2 \right\} = \mathrm{prox}_{\tau F}(-B\,z_k - \tau\,\lambda_k)$$

On démontre de même que, en posant $\nu_{k+1} = -B\,z_{k+1}$,

$$\nu_{k+1} \in \operatorname*{argmin}_{\nu \in \mathcal{Y}} \left\{ G(\nu) - \langle \lambda_k, \nu \rangle + \frac{1}{2\,\tau}\,\|A\,x_{k+1} - \nu - c\|^2 \right\} = \mathrm{prox}_{\tau G}(A\,x_{k+1} - c + \tau\,\lambda_k)$$

La mise-à-jour du multiplicateur de LAGRANGE s'écrit quant à elle

$$\tau\,\lambda_{k+1} = \tau\,\lambda_k + \xi_{k+1} - \nu_{k+1}$$

On en déduit d'une part que

$$-B\,z_k - \tau\,\lambda_k = \nu_k - \tau\,\lambda_k = \xi_k - \tau\,(2\,\lambda_k - \lambda_{k-1})$$

et, d'autre part, que $$\nu_{k+1} = (A\,x_{k+1} - c + \tau\,\lambda_k) - \tau\,\lambda_{k+1}$$

Utilisons l'identité de MOREAU (corollaire 5 du module **A8 : Dualité de FENCHEL et conjuguée convexe**) pour écrire que

$$A\,x_{k+1} - c + \tau\,\lambda_k = \mathrm{prox}_{\tau G}(A\,x_{k+1} - c + \tau\,\lambda_k) + \tau\,\mathrm{prox}_{G^*/\tau}\left(\frac{A\,x_{k+1} - c}{\tau} + \lambda_k\right)$$

Ainsi, la mise-à-jour de z_{k+1} équivaut à

$$\lambda_{k+1} = \mathrm{prox}_{G^*/\tau}\left(\lambda_k + \frac{1}{\tau}\,\xi_{k+1}\right)$$

et les itérations de l'ADMM s'écrivent

$$\begin{cases} \xi_{k+1} = \mathrm{prox}_{\tau F}(\xi_k - \tau\,(2\,\lambda_k - \lambda_{k-1})) \\ \lambda_{k+1} = \mathrm{prox}_{G^*/\tau}\left(\lambda_k + \dfrac{1}{\tau}\,\xi_{k+1}\right) \end{cases} \qquad \text{soit} \qquad \begin{cases} \xi_{k+1} = \mathrm{prox}_{\tau F}(\xi_k - \tau\,\tilde{\lambda}_k) \\ \lambda_{k+1} = \mathrm{prox}_{G^*/\tau}\left(\lambda_k + \dfrac{1}{\tau}\,\xi_{k+1}\right) \\ \tilde{\lambda}_{k+1} = 2\,\lambda_{k+1} - \lambda_k \end{cases}$$

On reconnaît les itérations de l'algorithme de CHAMBOLLE-POCK, appliqué au problème dual

$$\min_{\lambda \in \mathcal{Y}} \Big\{ F^*(-\lambda) + G^*(\lambda) \Big\} \qquad (\mathcal{D}_{\text{ADMM}})$$

dans le cas particulier (en utilisant les notations de la section précédente) où $f = F^*$, $g = G^*$, $A = -\mathrm{Id}$, $\theta = 1$ et $\sigma = 1/\tau$.

> De même que l'on a démontré que l'ADMM et l'algorithme de Chambolle–Pock sont équivalents, il est possible de montrer un lien entre l'algorithme de Douglas–Rachford (DRS) et l'ADMM.

Pour utiliser les résultats de la section précédente, on s'intéresse aux propriétés du problème $(\mathcal{D}_{\text{ADMM}})$. Notons tout d'abord qu'il s'agit d'un problème convexe si f et g le sont, et que les fonctions F et G sont s.c.i. Ensuite, montrons que, pour tout $\lambda \in \mathcal{Y}$,

$$F^*(-\lambda) = \sup_{\xi \in \mathcal{Y}} \left\{ -\langle \lambda, \xi \rangle - \inf_{\substack{x \in \mathcal{X} \\ A\,x - c = \xi}} f(x) \right\} = \sup_{\substack{(x,\xi) \in \mathcal{X} \times \mathcal{Y} \\ A\,x - c = \xi}} \Big\{ \langle \lambda, c - A\,x \rangle - f(x) \Big\}$$

Puisque ξ n'apparaît plus dans la dernière expression, on peut ignorer la restriction sur ξ, de sorte que, pour tout $\lambda \in \mathcal{Y}$,

$$F^*(-\lambda) = \sup_{x \in \mathcal{X}} \Big\{ \langle \lambda, c - A\,x \rangle - f(x) \Big\} = f^*(-A^*\,\lambda) + \langle \lambda, c \rangle$$

On montre de manière analogue que

$$\forall\, \lambda \in \mathcal{Y}, \qquad G^*(\lambda) = g^*(-B^*\lambda)$$

Ainsi, on a établi que

$$\forall\, \lambda \in \mathcal{Y}, \qquad F^*(-\lambda) + G^*(\lambda) = f^*(-A^*\,\lambda) + \langle \lambda, c \rangle + g^*(-B^*\lambda)$$

On reconnaît la fonction objectif du problème dual associé au lagrangien *ordinaire* du problème $(\mathcal{P}_{\text{ADMM}})$.

Lorsque l'on considère le problème dual sous la forme d'un problème de minimisation, il est possible de l'écrire comme un problème de recherche de point-selle, en écrivant que

$$\forall\, \lambda \in \mathcal{Y}, \qquad F^*(-\lambda) = \sup_{\xi \in \mathcal{Y}} \Big\{ \langle -\lambda, \xi \rangle - F(\xi) \Big\}$$

Aussi, on est amené à considérer la fonction de couplage

$$\tilde{\mathcal{L}}(\lambda; \xi) \mapsto G^*(\lambda) - \langle \lambda, \xi \rangle - F(\xi)$$

Comme on l'établira dans la section suivante, cette fonction de couplage présentant un saut de dualité nul. Il s'ensuit en particulier qu'elle admet un point-selle (λ^*, ξ^*), et que

$$\lambda^* \in \operatorname*{argmin}_{\lambda \in \mathcal{Y}} \Big\{ F^*(-\lambda) + G^*(\lambda) \Big\}$$

tandis que ξ^* minimise la fonction

$$\xi \mapsto \sup_{\lambda \in \mathcal{Y}} \Big\{ -G^*(\lambda) + \langle \lambda, \xi \rangle + F(\xi) \Big\} = G(\xi) + F(\xi)$$

L'absence de saut de dualité assure que la quantité $G(\xi^*) + F(\xi^*)$ est finie, donc qu'il existe $x^* \in \mathcal{X}$ et $z^* \in \mathcal{Z}$ tels que

$$A\,x^* - c = \xi^* \qquad \text{et} \qquad -B\,z^* = \xi^* \qquad \text{soit} \qquad A\,x^* + B\,z^* - c = 0$$

et tel que, pour tout $\xi \in \mathcal{Y}$

$$f(x^*) + g(z^*) = F(\xi^*) + G(\xi^*) \leq F(\xi) + G(\xi)$$

Or, par définition de F et de G, on a tout $(x,z) \in \mathcal{X} \times \mathcal{Z}$ tel que $A\,x + B\,z = c$

$$F(A\,x - c) + G(A\,x - c) \le f(x) + g(z)$$

On en déduit donc le point (x^*, z^*) est solution du problème initial $(\mathcal{P}_{\text{ADMM}})$. Notons que, même si ξ^* est unique, il n'en est pas nécessairement de même pour (x^*, z^*). Par ailleurs, pour déterminer ce point à partir de ξ^*, il faut résoudre les deux systèmes linéaires

$$A\,x^* = \xi^* + c \qquad \text{et} \qquad B\,z^* = -\xi^*$$

Commençons par traduire la proposition 3 pour l'ADMM (réécrit comme un cas particulier de l'algorithme de Chambolle–Pock) :

Proposition 4

Soit $f : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ et $g : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ deux fonctions **convexes**, s.c.i. et propres, $A : \mathcal{X} \to \mathcal{Y}$ et $B : \mathcal{Z} \to \mathcal{Y}$ deux opérateurs linéaire bornés et $c \in \mathcal{Y}$ un vecteur. Soit $((x_k, z_k, \lambda_k))_{k\in\mathbb{N}}$ la suite générée pour $(x_0, z_0, \lambda_0) \in \mathcal{X} \times \mathcal{Z} \times \mathcal{Y}$ par l'algorithme

$$\begin{cases} x_{k+1} \in & \operatorname*{argmin}_{x\in\mathcal{X}} \left\{ f(x) + \langle \lambda_k, A\,x\rangle + \dfrac{1}{2\,\tau}\,\|A\,x + B\,z_k - c\|^2 \right\} \\ z_{k+1} \in & \operatorname*{argmin}_{z\in\mathcal{Z}} \left\{ g(z) + \langle \lambda_k, B\,z\rangle + \dfrac{1}{2\,\tau}\,\|A\,x_{k+1} + B\,z - c\|^2 \right\} \\ \lambda_{k+1} = & \lambda_k + \dfrac{1}{\tau}\,(A\,x_{k+1} + B\,z_{k+1} - c) \end{cases}$$

Alors la suite $(\lambda_k)_{k\in\mathbb{N}}$ est bornée.

Démonstration : Il suffit de vérifier que $\tau \times (1/\tau) = 1 = 1/|||-\mathrm{Id}|||^2$ de sorte que les hypothèses des propositions 2 et 3 sont satisfaites. ■

On voit que ce résultat est relativement pauvre. Du fait que le paramètre de sur-relaxation θ vaut ici 1, et que le produit des pas de temps vaut 1 également, on ne peut pas espérer obtenir de résultats plus forts en utilisant les résultats de la section précédente.

Pour avoir des propriétés plus intéressantes, il faut considérer des problèmes plus réguliers. Si on suppose que f est $L_{\nabla f}$-régulière, et que g est fortement convexe de module α, alors, d'après le module **A8 : Dualité de Fenchel et conjuguée convexe**, f^* est fortement convexe de module $1/L_{\nabla f}$, tandis que g^* est $(1/\alpha)$-régulière. En particulier, $G^* = (g^*) \circ (-B^*)$ est $(|||B|||^2/\alpha)$-régulière, et donc si $|||B||| \neq 0$, alors G est fortement convexe, de module $(\alpha/|||B|||^2)$. Par ailleurs, la forte convexité de f^* permet d'écrire pour tout $t \in [\,0\,;1\,]$ et tout $(\lambda_1, \lambda_2) \in \mathcal{Y}^2$

$$\begin{aligned} F^*(t\,\lambda_1 + (1-t)\,\lambda_2) &\le t\,F^*(\lambda_1) + (1-t)\,F^*(\lambda_2) + \frac{1}{2\,L_{\nabla f}}\,t\,(1-t)\,\|A^*(\lambda_1 - \lambda_2)\|^2 \\ &\le t\,F^*(\lambda_1) + (1-t)\,F^*(\lambda_2) + \frac{|||A|||^2}{2\,L_{\nabla f}}\,t\,(1-t)\,\|\lambda_1 - \lambda_2\|^2 \end{aligned}$$

Autrement dit, si $|||A||| \neq 0$, alors F^* est fortement convexe, de module $|||A|||^2/L_{\nabla f}$ et F est donc $(L_{\nabla f}/|||A|||^2)$-régulière. On peut alors appliquer les résultats de convergence de l'algorithme de Chambolle–Pock dans le cas régulier, et démontrer la convergence de la suite des variables duales.

3.4 Exemple d'applications

Un exemple générique d'application de l'ADMM est le problème de la minimisation de la somme de deux fonctions simples :

$$\min_{x\in\mathcal{X}} \Big\{ f(x) + g(x) \Big\}$$

que l'on peut réécrire en dupliquant la variable x :

$$\min_{\substack{(x,z)\in\mathcal{X}\times\mathcal{Z}\\ x=z}} \Big\{ f(x) + g(z) \Big\}$$

On reconnaît un problème de la forme $(\mathcal{P}_{\text{ADMM}})$ avec $A = -B = \mathrm{Id}$ et $c = 0$. Les itérations de l'ADMM s'écrivent

$$\begin{cases} x_{k+1} = \operatorname*{argmin}_{x\in\mathcal{X}} \Big\{ f(x) + \langle \lambda_k, x\rangle + \dfrac{1}{2\,\tau}\,\|x - z_k\|^2 \Big\} \\ z_{k+1} = \operatorname*{argmin}_{z\in\mathcal{Z}} \Big\{ g(z) - \langle \lambda_k, z\rangle + \dfrac{1}{2\,\tau}\,\|x_{k+1} - z\|^2 \Big\} \\ \lambda_{k+1} = \lambda_k + \dfrac{1}{\tau}\,(x_{k+1} - z_{k+1}) \end{cases}$$

En factorisant les produits scalaires, on constate que ces itérations peuvent se réécrire

$$\begin{cases} x_{k+1} = \operatorname*{argmin}_{x\in\mathcal{X}} \Big\{ f(x) + \dfrac{1}{2\,\tau}\,\|x - z_k + \tau\,\lambda_k\|^2 \Big\} = \mathrm{prox}_{\tau f}(z_k - \tau\,\lambda_k) \\ z_{k+1} = \operatorname*{argmin}_{z\in\mathcal{Z}} \Big\{ g(z) + \dfrac{1}{2\,\tau}\,\|x_{k+1} - z + \tau\,\lambda_k\|^2 \Big\} = \mathrm{prox}_{\tau g}(x_{k+1} + \tau\,\lambda_k) \\ \lambda_{k+1} = \lambda_k + \dfrac{1}{\tau}\,(x_{k+1} - z_{k+1}) \end{cases}$$

On voit que le calcul des itérations ne nécessite que la simplicité de fonctions f et g. Il est important de noter à nouveau ici que l'identité $x_k = z_k$ n'est en général pas vérifiée avant convergence de l'algorithme.