Module B6

Éclatement primal-dual d'opérateurs

Sauf mention contraire, $\mathcal{X}$ est un espace vectoriel réel de dimension finie, muni du produit scalaire noté $\langle \cdot, \cdot \rangle$ et on note $\| \cdot \|$ la norme euclidienne associée. Pour toutes les autres notations et définitions, le lecteur est invité à se reporter aux modules précédents.

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 $\mathcal{L} : \mathcal{X} \times \mathcal{Y} \to \mathbb{R} \cup \{+\infty, -\infty\}$ vérifiant

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

de sorte que, par dualité min-max (voir module A6), 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

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

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 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 déjà abordées dans les modules précédents, 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 trois exemples de fonction de couplage, à savoir le lagrangien ordinaire, le lagrangien augmenté et l'utilisation de la conjuguée convexe.

2 Méthode d'Uzawa

2.1 Position du problème

On s'intéresse aux problèmes d'optimisation sous contraintes d'inégalités

$$\min_{\substack{x \in \mathcal{X} \\ Ax - b \leq 0}} J(x) \tag{$\mathcal{P}_{\text{Uzawa}}$}$$

avec $J : \mathcal{X} \to \mathbb{R} \cup \{+\infty\}$ une fonction **convexe**, propre, s.c.i. et fortement convexe de module $\alpha > 0$, $A : \mathcal{X} \to \mathcal{Y}$ un opérateur linéaire borné sur $\mathcal{X}$ et $b \in \mathcal{Y}$. On introduit alors le lagrangien ordinaire associé à ce problème, donné par :

$$\forall\, (x, \lambda) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(x; \lambda) = \begin{cases} J(x) + \langle \lambda, A\,x - b \rangle & \text{si } \lambda \in (\mathbb{R}^+)^m \\ -\infty & \text{sinon} \end{cases}$$

La fonction duale associée est la fonction concave définie par

$$E(\lambda) = \inf_{x \in \mathcal{X}} \mathcal{L}(x; \lambda) = \begin{cases} \inf_{x \in \mathcal{X}} \big\{ J(x) + \langle \lambda, A\,x - b \rangle \big\} & \text{si } \lambda \in (\mathbb{R}^+)^m \\ -\infty & \text{sinon} \end{cases}$$

et le problème dual est donc donné par

$$\max_{\lambda \in (\mathbb{R}^+)^m} E(\lambda) \tag{$\mathcal{D}_{\text{Uzawa}}$}$$

Pour résoudre le problème $(\mathcal{P}_{\text{Uzawa}})$, on a vu dans le module A7 qu'il est possible de rechercher les points-selles du lagrangien, c'est-à-dire résoudre le problème

$$\min_{x \in \mathcal{X}} \sup_{\lambda \in (\mathbb{R}^+)^m} \big\{ J(x) + \langle \lambda, A\,x - b \rangle \big\}$$

Autrement dit, on recherche un point $(\bar{x}, \bar{\lambda}) \in \mathcal{X} \times (\mathbb{R}^+)^m$ tel que $\bar{x}$ soit un minimiseur de la fonction partielle $x \mapsto \mathcal{L}(x; \bar{\lambda})$ et $\bar{\lambda}$ un maximiseur de la fonction partielle $\mathcal{L}(\bar{x}; \lambda)$.

2.2 Algorithme

C'est l'objet de l'algorithme d'Uzawa, introduit en 1958 par Hirofumi Uzawa. Les itérations de cet algorithme sont données par

$$(x_0, \lambda_0) \in \mathcal{X} \times (\mathbb{R}^+)^m \qquad \text{et} \qquad \forall\, k \in \mathbb{N}, \quad \begin{cases} x_{k+1} = & \operatorname*{argmin}_{x \in \mathcal{X}} \mathcal{L}(x; \lambda_k) \\ \lambda_{k+1} = & \operatorname{proj}_{(\mathbb{R}^+)^m} \left(\lambda_k + \dfrac{1}{\tau} \dfrac{\partial \mathcal{L}}{\partial \lambda}(x_{k+1}; \lambda_k) \right) \end{cases}$$

Elles peuvent être interprétées de la manière suivante : la méthode d'Uzawa alterne

1. une **minimisation** partielle en x, la variable primale ;
2. une **montée de gradient projeté** en λ, la variable duale.

Un tel choix est aisé à comprendre, puisque le point recherché $\bar{x}$ est un minimiseur et le multiplicateur associé $\bar{\lambda}$ est un minimiseur de la fonction convexe $\lambda \mapsto -\mathcal{L}(\bar{x};\lambda)$, pour la recherche duquel la mise en place d'un schéma de descente de gradient projeté est naturelle.

On notera par ailleurs que, dans les itérations d'Uzawa, les problèmes de minimisation sont non contraints (donc, *a priori* plus simples que le problème initial $(\mathcal{P}_{\text{Uzawa}})$). Plus précisément, il s'agit de problèmes fortement convexes, ce qui implique l'existence et l'unicité du minimiseur. De plus, les projections sont simples à calculer, puisque

$$\forall\, \lambda = (\lambda_{(j)})_{1\leq j\leq m}, \qquad \operatorname{proj}_{(\mathbb{R}^+)^m}(\lambda) = \big(\max(0,\lambda_{(j)})\big)_{1\leq j\leq m}$$

L'intérêt de cette méthode se comprend donc quand la projection sur l'ensemble admissible $\{x \in \mathcal{X} \mid A\,x - b \leq 0\}$ n'est pas calculable (car, autrement, et si J est différentiable, on pourrait appliquer la méthode du gradient projeté).

2.3 Propriétés de convergence

Commençons par noter que le problème dual (sous forme d'un problème de minimisation) s'écrit à l'aide de la conjuguée convexe J^* de J :

$$\min_{\lambda\in(\mathbb{R}^+)^m} \Big\{ J^*(-A^*\,\lambda) + \langle \lambda, b\rangle \Big\}$$

La k-ème itération de la méthode d'Uzawa est alors donnée par

$$\begin{cases} x_{k+1} = & \operatorname*{argmin}_{x\in\mathcal{X}} \Big\{ J(x) + \langle \lambda_k, A\,x - b\rangle \Big\} \\ \lambda_{k+1} = & \operatorname{proj}_{\lambda\in(\mathbb{R}^+)^m} \Big(\lambda_k + \dfrac{1}{\tau}\,(A\,x_{k+1} - b) \Big) \end{cases}$$

Écrivons la CNS d'optimalité du premier ordre pour la mise-à-jour de la variable primale ; en utilisant la règle de bascule, on obtient que

$$-A^*\lambda_k \in \partial J(x_{k+1}) \qquad \Longleftrightarrow \qquad x_{k+1} = \nabla(J^*)(-A^*\lambda_k)$$

Aussi, les itérations de l'algorithme d'Uzawa s'écrivent

$$\lambda_0 \in (\mathbb{R}^+)^m \qquad \text{et} \qquad \lambda_{k+1} = \operatorname{proj}_{(\mathbb{R}^+)^m} \Big(\lambda_k - \frac{1}{\tau}\,(-A\,\nabla J^*(-A^*\lambda_k) + b) \Big)$$

On reconnaît les itérations de l'algorithme de gradient projeté appliqué au problème dual, de pas de temps $(1/\tau)$. Or, en utilisant la proposition 13 du module A8, on peut montrer J^* est différentiable, de gradient lipschitzien de constante $1/\alpha$. On suppose que A est borné. Alors, la fonction $-E$ est également régulière, puisque

$$\begin{aligned} \|\nabla(-E)(\lambda) - \nabla(-E)(\lambda')\| &= \| - A\,\nabla(J^*)(-A^*\lambda) + b - A\,\nabla(J^*)(-A^*\lambda') - b\| \\ &\leq |||A|||\, \|\nabla(J^*)(-A^*\lambda) - \nabla(J^*)(-A^*\lambda')\| \\ &\leq \frac{|||A|||}{\alpha}\, \| - A^*\lambda + A^*\lambda'\| \\ \|\nabla(-E)(\lambda) - \nabla(-E)(\lambda')\| &\leq \frac{|||A|||^2}{\alpha}\, \|\lambda - \lambda'\| \end{aligned}$$

On peut alors établir le résultat suivant :

Lemme 1

Soit $J : \mathcal{X} \mapsto \mathbb{R}\cup\{+\infty\}$ est une fonction fortement **convexe** de module α, $A : \mathcal{X} \mapsto \mathcal{Y}$ un opérateur linéaire borné et $b \in \mathcal{Y}$. On pose

$$\forall\, \lambda \in \mathcal{Y}, \qquad E(\lambda) = -J^*(-A^*\,\lambda) - \langle \lambda, b\rangle$$

et on suppose que $-E$ est minorée sur $(\mathbb{R}^+)^m$. Soit $(\lambda_k)_{k\in\mathbb{N}}$ une suite générée par l'algorithme suivant : soit $\lambda_0 \in (\mathbb{R}^+)^m$ et

$$\forall\, k \in \mathbb{N}, \quad \lambda_{k+1} = \operatorname{proj}_{(\mathbb{R}^+)^m} \Big(\lambda_k - \frac{1}{\tau}\,(-A\,\nabla J^*(-A^*\lambda_k) + b) \Big)$$

Si $\tau \geq |||A|||^2/\alpha$, alors la suite $(\lambda_k)_{k\in\mathbb{N}}$ converge vers un minimiseur de E.

Démonstration : Il s'agit d'une conséquence de la proposition 7 du module B4.

Notons $\lambda^* \in (\mathbb{R}^+)^m$ la limite des λ_k. Sous les hypothèses de la proposition précédente, on peut utiliser la continuité de $\nabla(J^*)$ pour assurer que

$$\lim_{k\to+\infty} x_{k+1} = \lim_{k\to+\infty} \nabla(J^*)(-A^*\lambda_k) = \nabla(J^*)(-A^*\lambda^*)$$

de sorte que, si on note x^* la limite des x_k, on a grâce à la règle de bascule

$$-A^*\lambda^* \in \partial J(x^*)$$

Autrement dit, x^* est un minimiseur de la fonction partielle $x \mapsto \mathcal{L}(x;\lambda^*)$, c'est-à-dire que

Proposition 1

Soit $J : \mathcal{X} \mapsto \mathbb{R}\cup\{+\infty\}$ est une fonction fortement **convexe** de module α, $A : \mathcal{X} \mapsto \mathcal{Y}$ un opérateur linéaire borné et $b \in \mathcal{Y}$. On pose

$$\forall\, \lambda \in \mathcal{Y}, \qquad E(\lambda) = -J^*(-A^*\,\lambda) - \langle \lambda, b\rangle$$

et on suppose que $-E$ est minorée sur $(\mathbb{R}^+)^m$. Soit $((x_k, \lambda_k))_{k\in\mathbb{N}}$ une suite générée par l'algorithme suivant : soit $\lambda_0 \in (\mathbb{R}^+)^m$ et

$$\forall\, k \in \mathbb{N}, \quad \begin{cases} x_{k+1} = & \operatorname*{argmin}_{x\in\mathcal{X}} \Big\{ J(x) + \langle \lambda_k, A\,x - b\rangle \Big\} \\ \lambda_{k+1} = & \operatorname{proj}_{(\mathbb{R}^+)^m} \Big(\lambda_k + \dfrac{1}{\tau}\,(A\,x_{k+1} - b) \Big) \end{cases}$$

Si $\tau \geq |||A|||^2/\alpha$, alors la suite $(x_k)_{k\in\mathbb{N}}$ converge vers un minimiseur de J.

2.4 Exemple d'applications

On suppose que J est une fonction quadratique généralisée

$$\forall\, x \in \mathcal{X}, \qquad J(x) = \frac{1}{2}\,\|A\,x\|^2 + \langle \beta, x\rangle + \gamma$$

avec $K : \mathcal{X} \to \mathcal{Z}$ une application linéaire, $\beta \in \mathcal{X}$ et $\gamma \in \mathbb{R}$. On considère alors le problème d'optimisation sous contraintes

$$\min_{\substack{x\in\mathcal{X}\\ A\,x-b\leq 0}} J(x)$$

avec $A : \mathcal{X} \to \mathcal{Y}$ une application linéaire et $b \in \mathcal{Y}$. On pourra noter que, sans perte de généralité, on peut ignorer la constante γ, c'est-à-dire supposer que $\gamma = 0$. Le lagrangien de ce problème est donné par

$$\forall\,(x,\lambda) \in \mathcal{X} \times (\mathbb{R}^+)^m, \qquad \mathcal{L}(x;\lambda) = \frac{1}{2}\,\|A\,x\|^2 - \langle \beta, x\rangle + \langle \lambda, A\,x - b\rangle$$

Ses points-selles sont les points $(\bar{x}, \bar{\lambda})$ vérifiant

$$\frac{\partial \mathcal{L}}{\partial x}(\bar{x}, \bar{\lambda}) = K^* A\,\bar{x} - \beta + A^*\,\bar{\lambda} = 0 \qquad \text{et} \qquad \frac{\partial \mathcal{L}}{\partial \lambda}(\bar{x}, \bar{\lambda}) = A\,\bar{x} - b = 0$$

Autrement dit, il s'agit des solutions du système linéaire

$$\underbrace{\begin{pmatrix} K^*K & A^* \\ A & 0 \end{pmatrix}}_{=M} \begin{pmatrix} x \\ \lambda \end{pmatrix} = \begin{pmatrix} \beta \\ b \end{pmatrix}$$

où la matrice M est une matrice symétrique, semi-définie positive. La méthode d'Uzawa consiste alors à calculer les itérations suivantes :

$$(x_0, \lambda_0) \in \mathcal{X} \times (\mathbb{R}^+)^m \qquad \text{et} \qquad \forall\,k \in \mathbb{N}, \quad \begin{cases} x_{k+1} = & (K^*K)^{-1}(\beta - A^*\lambda_k) \\ \lambda_{k+1} = & \operatorname{proj}_{(\mathbb{R}^+)^m}\left(\lambda_k + \dfrac{1}{\tau}\,(A\,x_k - b)\right) \end{cases}$$

On voit donc qu'on a remplacé la résolution d'un grand système linéaire (de taille $d \times d$) par une succession de résolutions de systèmes linéaires plus petits (de taille $(d-m)\times(d-m)$), la mise-à-jour des multiplicateurs de Lagrange étant de complexité négligeable. Il apparaît donc que l'algorithme d'Uzawa est, dans ce cas, avantageux lorsque m est grand devant la dimension de $\mathcal{X}$.

Pour pouvoir appliquer les résultats de convergence, on vérifie que la fonction objectif J est fortement convexe. Pour cela, il suffit de calculer son gradient, qui vaut

$$\forall\,x \in \mathcal{X}, \qquad \nabla J(x) = K^* A\,x + \beta$$

qui vérifie

$$\forall\,(x,x') \in \mathcal{X}^2, \qquad \langle \nabla J(x) - \nabla J(x'), x - x'\rangle = \|A\,x - A\,x'\|^2 \geq |||A|||^2\,\|x - x'\|^2$$

ce qui prouve que le module de forte convexité de J vaut $|||A|||^2$.

3 Méthode des directions alternées

3.1 Position du problème

On s'intéresse aux problèmes de la forme

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

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

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

Pour tout $(x, z) \in \operatorname{dom} f \times \operatorname{dom} g = \operatorname{dom} J$, la fonction $\lambda \mapsto \mathcal{L}_\tau(x,z;\lambda)$ est différentiable, de gradient

$$\frac{\partial \mathcal{L}_\tau}{\partial \lambda}(x,z;\lambda) = A\,x + B\,z - c$$

qui est lipschitzien de constante 0. Le problème ($\mathcal{P}_{\text{ADMM}}$) s'écrit 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

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 :

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

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

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

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 module B5) appliqué à ce problème de minimisation.

3.3 Propriétés de convergence I

Dans ce paragraphe, on démontre qu'il existe un lien entre l'ADMM et la méthode de Douglas–Rachford (module B4), de sorte que l'on pourra utiliser les résultats de convergence de cette dernière méthode.

Commençons par introduire le lagrangien *ordinaire* associé au problème ($\mathcal{P}_{\text{ADMM}}$) :

$$\forall\,(x, z, \lambda) \in \mathcal{X}\times\mathcal{Z}\times\mathcal{Y}, \qquad \mathcal{L}(x, z; \lambda) = f(x) + g(z) + \langle \lambda, A\,x + B\,z - c\rangle$$

Puisque, pour tout $\lambda \in \mathcal{Y}$, on a

$$\begin{aligned} E(\lambda) &= \inf_{(x,z)\in\mathcal{X}\times\mathcal{Z}} \mathcal{L}(x, z; \lambda) \\ &= \inf_{x\in\mathcal{X}} \Big\{ f(x) + \langle \lambda, A\,x\rangle \Big\} + \inf_{z\in\mathcal{Z}} \Big\{ g(z) + \langle \lambda, B\,z\rangle \Big\} - \langle \lambda, c\rangle \\ &= -\sup_{x\in\mathcal{X}} \Big\{ \langle -A^*\lambda, x\rangle - f(x) \Big\} - \sup_{z\in\mathcal{Z}} \Big\{ \langle -B^*\lambda, z\rangle - g(z) \Big\} - \langle \lambda, c\rangle \\ E(\lambda) &= -f^*(-A^*\lambda) - g^*(-B^*\lambda) - \langle \lambda, c\rangle \end{aligned}$$

on en déduit que le problème dual prend la forme

$$\min_{\lambda\in\mathcal{Y}} \Big\{ f^*(-A^*\lambda) + g^*(-B^*\lambda) + \langle \lambda, c\rangle \Big\}$$

On pose

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

Alors leur sous-différentiel respectif en $\lambda \in \mathcal{Y}$ vaut :

$$\partial F(\lambda) = -B\,(\partial(g^*)(-B^*\,\lambda)) \qquad \text{et} \qquad \partial G(\lambda) = -A\,(\partial(f^*)(-A^*\,\lambda)) + c$$

On commence par s'intéresser à la définition de z_{k+1}. Ce point est caractérisé par la CNS d'optimalité du premier ordre, qui s'écrit

$$-B^*\left(\lambda_k + \frac{1}{\tau}\,(A\,x_{k+1} + B\,z_{k+1} - c)\right) \in \partial g(z_{k+1})$$

Appliquons la règle de bascule sur cette relation ; on obtient alors que

$$z_{k+1} \in \partial(g^*)\left(-B^*\left(\lambda_k + \frac{1}{\tau}\,(A\,x_{k+1} + B\,z_{k+1} - c)\right)\right) = \partial(g^*)\,(-B^*\lambda_{k+1})$$

soit $$-B\,z_{k+1} \in -B\,(\partial(g^*)\,(-B^*\lambda_{k+1})) = \partial F\,(\lambda_{k+1})$$

En utilisant le fait que

$$\tau\,(x^0 - x^+) \in \partial F(x^+) \qquad \Longleftrightarrow \qquad x^+ = \operatorname{prox}_{F/\tau}(x^0)$$

on en déduit finalement que

$$\lambda_{k+1} = \operatorname{prox}_{F/\tau}(p_k) \qquad \text{avec} \qquad \forall\,k \in \mathbb{N}, \quad p_k = \lambda_{k+1} - \frac{1}{\tau}\,B\,z_{k+1}$$

On s'intéresse maintenant à la mise-à-jour des x_k. À l'itération $k+2$, on a, par la règle de Fermat,

$$-A^*\left(\lambda_{k+1} + \frac{1}{\tau}\,(A\,x_{k+2} + B\,z_{k+1} - c)\right) \in \partial f(x_{k+2})$$

En utilisant à nouveau la règle de bascule, puis en composant l'inclusion obtenue par $p \mapsto -A\,p + c$, on obtient que

$$-A\,x_{k+2} + c \in \partial G\left(\lambda_{k+1} + \frac{1}{\tau}\,(A\,x_{k+2} + B\,z_{k+1} - c)\right)$$

Posons pour tout $k \in \mathbb{N}$

$$p_{k+1/2} = \lambda_{k+1} + \frac{1}{\tau}\,(A\,x_{k+2} + B\,z_{k+1} - c)$$

En utilisant la caractérisation du point proximal, on obtient alors que

$$p_{k+1/2} = \operatorname{prox}_{G/\tau}\left(\frac{1}{\tau}\,(A\,x_{k+2} - c) - p_{k+1/2}\right) = \operatorname{prox}_{G/\tau}\,(p_k - 2\,\lambda_{k+1})$$

Par ailleurs, on remarque que

$$\begin{aligned} p_{k+1} - p_k &= \lambda_{k+2} - \lambda_{k+1} - \frac{1}{\tau}\,B\,(z_{k+2} - z_{k+1}) \\ &= \frac{1}{\tau}\,(A\,x_{k+2} + B\,z_{k+1} - c) \\ p_{k+1} - p_k &= p_{k+1/2} - \lambda_{k+1} \end{aligned}$$

Finalement, on a réécrit les itérations de l'ADMM sous la forme

$$\begin{cases} \lambda_{k+1} &= \operatorname{prox}_{F/\tau}(p_k) \\ p_{k+1/2} &= \operatorname{prox}_{G/\tau}\,(p_k - 2\,\lambda_{k+1}) \\ p_{k+1} &= p_k + p_{k+1/2} - \lambda_{k+1} \end{cases}$$

On reconnaît les itérations de la méthode de Douglas–Rachford (module B4) avec, en utilisant les notations de la section 2 du module B4, $\mu = 1$, appliqué au problème dual

$$\min_{\lambda\in\mathcal{Y}} \Big\{ F(\lambda) + G(\lambda) \Big\}$$

Il est aisé de vérifier que les fonctions F et G sont convexes, s.c.i. et propres. Aussi, si on a de plus la condition

$$0 \in \partial F(\lambda^*) + \partial G(\lambda^*)$$

pour un certain $\lambda^* \in \mathcal{Y}$, la proposition 12 du module B4 assure que

Lemme 2

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. On suppose qu'il existe $\lambda^* \in \mathcal{Y}$ tel que

$$0 \in -A\,(\partial(f^*)(-A^*\,\lambda^*)) - B\,(\partial(g^*)(-B^*\,\lambda^*)) + c$$

On pose

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

Soit $((\lambda_k, p_{k+1/2}, p_k))_{k\in\mathbb{N}}$ la suite générée par l'algorithme

$$\begin{cases} \lambda_{k+1} & = \operatorname{prox}_{F/\tau}(p_k) \\ p_{k+1/2} & = \operatorname{prox}_{G/\tau}(p_k - 2\,\lambda_{k+1}) \\ p_{k+1} & = p_k + p_{k+1/2} - \lambda_{k+1} \end{cases}$$

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

$$\lim_{k\to+\infty} \|\lambda_{k+1} - \lambda_k\| = 0$$

Démonstration : Il suffit d'appliquer la propositon 12 du module B4. On démontre en particulier que la suite $(p_k)_{k\in\mathbb{N}}$ est bornée et que

$$\lim_{k\to+\infty} \|p_{k+1} - p_k\| = 0$$

On peut alors utiliser le caractère lipschitzien de l'opérateur proximal. ■

Proposition 2

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. De plus,

$$\lim_{k\to+\infty} \|\lambda_{k+1} - \lambda_k\| = 0$$

3.4 Propriétés de convergence II

On va maintenant démontrer qu'il existe un lien entre l'ADMM et un autre algorithme, celui de Chambolle–Pock, que l'on étudiera à la prochaine section. Rappelons l'expression des itérations de l'ADMM :

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

Considérons les deux fonctions

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

et

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

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

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

En particulier, on a

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

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

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

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 $\qquad \nu_{k+1} = (A\,x_{k+1} - c + \tau\,\lambda_k) - \tau\,\lambda_{k+1}$

Utilisons l'identité de MOREAU (corollaire 5 du module A8) pour écrire que

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

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

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

et les itérations de l'ADMM s'écrivent

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

On reconnaît les itérations de l'algorithme de CHAMBOLLE-POCK que l'on présentera à la section suivante, appliqué au problème dual

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

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

Pour utiliser les résultats de la section suivante, 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 $\lambda \in \mathcal{Y}$,

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

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

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

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

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

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

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

tandis que ξ^* minimise la fonction

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

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 $A\,x + B\,z = c$

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

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 7 pour l'ADMM (réécrit comme un cas particulier de l'algorithme de CHAMBOLLE–POCK) :

Proposition 3

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 6 et 7 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.

On suppose donc que f est $L_{\nabla f}$-régulière, et que g est fortement convexe de module α. Il s'ensuit que f^* est fortement convexe de module $1/L_{\nabla_f}$-régulière, et 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 $(\lambda_1, \lambda_2) \in \mathcal{Y}^2$

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

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 donc appliquer les résultats de la section suivante relativement au cas dit régulier. En particulier, la proposition 8 s'écrit :

Proposition 4 (Convergence duale dans le cas régulier)

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. On suppose que f est $L_{\nabla f}$-régulière, et que g est fortement convexe de module α. 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

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

où $\tau > 0$. Si $\tau = |||A||| \cdot |||B|||/\sqrt{\alpha\,L_{\nabla_f}}$ et ω tel que

$$\frac{1}{2 + |||A|||\,\sqrt{\alpha/L_{\nabla_f}}} \leq \omega \leq 1$$

alors la suite $(\lambda_k)_{k\in\mathbb{N}}$ converge vers l'unique maximiseur λ^* de la fonction

$$E : \lambda \mapsto f^*(-A^*\,\lambda) + \langle \lambda, c\rangle + g^*(-B^*\,\lambda)$$

et il existe $C > 0$ tel que

$$\forall\,k \in \mathbb{N}^*, \qquad \|\lambda^* - \lambda_k\|^2 \leq \omega^k\,C$$

Par ailleurs, on a

$$\lim_{k\to+\infty} E(\lambda_k) = \min_{\substack{(x,z)\in\mathcal{X}\times\mathcal{Z}\\ Ax+Bz=c}} \left\{ f(x) + g(z) \right\}$$

De plus, les suites $(A\,x_k)_{k\in\mathbb{N}}$ et $(B\,z_k)_{k\in\mathbb{N}}$ convergent également, et la somme de leur limite vaut c.

Démonstration : La dernière partie de la proposition est une conséquence de la convergence de la suite des $\xi_k = A\,x_k$, qui implique, grâce à l'égalité

$$B\,z_k = \tau\,(\lambda_k - \lambda_{k-1}) - A\,x_k + c$$

que la suite des $B\,z_k$ converge également. La convergence du critère dual est une conséquence de la continuité de la fonction F, elle-même conséquence de la continuité de f (on le démontre par l'absurde). ■

Évidemment, si A et B sont injectifs, alors on en déduit directement la convergence des variables primales $((x_k, z_k))_{k\in\mathbb{N}}$. Dans le cas général, il n'est pas possible d'assurer leur convergence. Néanmoins, on peut vérifier que toute valeur d'adhérence de cette suite est un minimiseur du problème $(\mathcal{P}_{\text{ADMM}})$.

3.5 Exemple d'applications

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

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

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

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

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

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

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

$$\left\{\begin{aligned} x_{k+1} &= \operatorname*{argmin}_{x\in\mathcal{X}} \left\{ f(x) + \frac{1}{2\,\tau}\,\|x - z_k + \tau\,\lambda_k\|^2 \right\} = \operatorname{prox}_{\tau f}(z_k - \tau\,\lambda_k) \\ z_{k+1} &= \operatorname*{argmin}_{z\in\mathcal{Z}} \left\{ g(z) + \frac{1}{2\,\tau}\,\|x_{k+1} - z + \tau\,\lambda_k\|^2 \right\} = \operatorname{prox}_{\tau g}(x_{k+1} + \tau\,\lambda_k) \\ \lambda_{k+1} &= \lambda_k + \frac{1}{\tau}\,(x_{k+1} - z_{k+1}) \end{aligned}\right.$$

On voit que le calcul des itérations ne nécessite que la simplicité de fonctions f et g.

4 Algorithme de Chambolle–Pock

4.1 Position du problème

On s'intéresse aux problèmes de la forme

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

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

$$\forall\, x \in \mathcal{X}, \qquad f(A\,x) = \sup_{y\in\mathcal{Y}} \Big\{ \langle y, A\,x\rangle - f^*(y) \Big\}$$

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

$$\min_{x\in\mathcal{X}} \sup_{y\in\mathcal{Y}} \Big\{ g(x) + \langle y, A\,x\rangle - f^*(y) \Big\}$$

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

Intéressons-nous au problème dual associé

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

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

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

Cette fonction étant concave, la CNS d'optimalité (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 CNS d'optimalité pour le problème primal : x^* est solution de ce problème si et seulement si

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

On suppose qu'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 tout d'abord que (identité de Legendre–Fenchel, lemme 3 du module A8)

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

En sommant ses deux identités, on en déduit que

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

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

Proposition 5

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 affirmations suivantes :

(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

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

4.2 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***).

4.3 Propriétés de convergence

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

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

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 CNS d'optimalité du premier ordre 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 3

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 :

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

$$\forall\, x \in \mathcal{X}, \qquad g(x) \geq g(x^+) + \frac{1}{\tau}\,\langle \bar{x} - x^+, x - x^+ \rangle - \langle \tilde{y}, A\,(x - x^+) \rangle$$

En ajoutant respectivement

$$\frac{1}{2\,\sigma}\,\|y - \bar{y}\|^2 \qquad \text{et} \qquad \frac{1}{2\,\tau}\,\|x - \bar{x}\|^2$$

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 4 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 4

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 τ, $\sigma > 0$. 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 $\qquad \forall k \in \mathbb{N}, \qquad \Delta_k = \dfrac{1}{2\,\sigma}\,\|y-y_k\|^2 + \dfrac{1}{2\,\tau}\,\|x-x_k\|^2$

Lemme 5 (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 $\tau,\, \sigma > 0$. 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, l'inégalité

de Jensen assure que 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 5 s'écrit

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

Proposition 6 (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 $|||A|||$. On pose

$$\forall (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(x; y) = g(x) + \langle y, A\,x\rangle - f^*(y)$$

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 $\tau,\, \sigma > 0$. 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 5 à un point-selle $(x, y) = (\hat{x}, \hat{y})$ 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 7 (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 $|||A|||$. On pose

$$\forall\, (x, y) \in \mathcal{X} \times \mathcal{Y}, \qquad \mathcal{L}(x; y) = g(x) + \langle y, A\, x\rangle - f^*(y)$$

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 $\tau\,\sigma \le 1/|||A|||^2$, 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

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

Démonstration : Décomposons la preuve en plusieurs étapes.

- **La suite $(x_k)_{k\in\mathbb{N}}$ est bornée.** On reprend l'inégalité du lemme 5 ; on a pour tout $K \in \mathbb{N}^*$

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

car $|||A|||\, \sqrt{\tau\,\sigma} \le 1$ 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 5 que, pour tout $K \in \mathbb{N}^*$,

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

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.,

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

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 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 $j \ge j_0$,

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

Soit $K \ge k_{j_0}$. En appliquant le lemme 5 aux indices K et k_{j_0}, on obtient après soustraction que

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

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

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

Il s'ensuit que

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

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

On s'intéresse maintenant au cas où la fonction g fortement convexe de module α et la fonction f $L_{\nabla f}$-régulière. Il est aisé de vérifier que $f \circ A$ est dans ce cas $|||A|||^2\, L_{\nabla f}$-régulière. Posons

$$J : x \mapsto f(A\, x) + g(x)$$

On définit son *conditionnement* comme la quantité positive

$$\kappa = \frac{|||A|||^2\, L_{\nabla f}}{\alpha}$$

Cette définition est à rapprocher de celle du conditionnement d'une fonction régulière fortement convexe. En particulier, sous les hypothèses faites plus haut, J est fortement convexe de module α ; aussi, dès que g est régulière, la quantité κ minore le conditionnement de J. On notera que, dans le cas général que l'on considère ici, le conditionnement peut être inférieur à 1.

Avant d'établir des résultats de convergence pour l'algorithme de Chambolle–Pock, notons que, si (x^*, y^*) est un point-selle de $\mathcal{L}$, alors

$$\forall\, k \in \mathbb{N}, \qquad 0 \in \partial_x \mathcal{L}(x^*; y_{k+1}) \qquad \text{et} \qquad 0 \in \partial_y(-\mathcal{L})(x_{k+1}; y^*)$$

La forte convexité de g assure alors que

$$g(x_{k+1}) + \langle y_{k+1}, A\, x_{k+1}\rangle \ge g(x^*) + \langle y_{k+1}, A\, x^*\rangle + \frac{\alpha}{2}\, \|x^* - x_{k+1}\|^2$$

tandis que celle de f^* (qui découle de la régularité de f) permet d'écrire

$$f^*(y_{k+1}) - \langle y_{k+1}, A\,x_{k+1}\rangle \geq f^*(y^*) - \langle y^*, A\,x_{k+1}\rangle + \frac{1}{2\,L_{\nabla_f}}\,\|y^* - y_{k+1}\|^2$$

Sommons ces deux inégalités ; on obtient

$$\mathcal{L}(x_{k+1}; y^*) - \mathcal{L}(x^*; y_{k+1}) \geq \frac{\alpha}{2}\,\|x^* - x_{k+1}\|^2 + \frac{1}{2\,L_{\nabla_f}}\,\|y^* - y_{k+1}\|^2$$

Revenons aux propriétés de convergence de l'algorithme de Chambolle–Pock. Une variante du lemme 3 est le lemme suivant :

Lemme 6 (Cas régulier)

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. On suppose que g est fortement convexe, de module α et que f est $L_{\nabla f}$-régulière. 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

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

Démonstration : La forte convexité de g assure que

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

tandis que, d'après la dualité forte convexité /régularité, la forte convexité de f^*, de module $1/L_{\nabla f}$, permet d'écrire

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

et on reprend la preuve du lemme 3. ■

En utilisant le lemme 4, on peut démontrer une variante du lemme 5 :

Lemme 7 (Cas régulier, $\theta < 1$)

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 $|||A|||$. On suppose que g est fortement convexe, de module α et que f est $L_{\nabla f}$-régulière. 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

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

où τ, $\sigma > 0$ et $\theta \in\,]\,0\,;1\,[$. Si $\sigma = \tau\,\alpha\,L_{\nabla_f}$, $\tau\,\sigma\,|||A|||^2 \leq 1$ et $\theta \geq 1/(1 + \tau\,\alpha)$, alors on a pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$

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

pour tout ω tel que $\qquad \dfrac{1 + \theta}{2 + \tau\,\alpha} \leq \omega \leq \theta$

Démonstration : On combine cette fois les résultats des lemmes 6 et 4, en posant $\mu = \sigma/L_{\nabla f} = \tau\,\alpha$:

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

Si $1 \geq |||A|||\sqrt{\tau\,\sigma}$, alors $1 \geq \theta \geq \theta\,|||A|||\sqrt{\tau\,\sigma}$ et

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

On choisit ω et θ tel que

$$\omega\,(1 + \mu) - 1 + \omega - \theta \geq 0 \qquad \text{soit} \qquad \omega \geq \frac{1 + \theta}{2 + \mu}$$

de sorte que les deux derniers termes sont positifs. On notera qu'un tel choix est possible, car

$$\theta \geq \frac{1 + \theta}{2 + \mu} \qquad \Longleftrightarrow \qquad \theta\,(1 + \mu) \geq 1$$

On divise alors par ω^k et on somme pour k entre 0 et $K-1$:

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

Après télescopage des termes, on obtient :

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

On minore le produit scalaire à l'aide de $\beta > 0$:

$$-\langle y - y_K, A\,(x_K - x_{K-1})\rangle \geq -\frac{\beta\,|||A|||}{2}\,\|y-y_K\|^2 - \frac{|||A|||}{2\,\beta}\,\|x_K - x_{K-1}\|^2$$

et on voit qu'il suffit de choisir $\beta = \tau\,|||A|||$ pour que les termes en $\|x_K - x_{K-1}\|^2$ s'annulent ; ainsi, on obtient le résultat annoncé. ■

Considérons cette fois-ci pour tout $K \in \mathbb{N}^*$ les moyennes pondérées

$$\tilde{X}_K = \frac{1}{T_K}\sum_{k=0}^{K-1}\frac{x_{k+1}}{\omega} \qquad \text{et} \qquad \tilde{Y}_K = \frac{1}{T_K}\sum_{k=0}^{K-1}\frac{y_{k+1}}{\omega} \qquad \text{avec} \qquad T_K = \sum_{k=0}^{K-1}\frac{1}{\omega}$$

Ainsi, en utilisant l'inégalité de Jensen, le lemme 7 devient

$$\begin{aligned}\frac{1}{2\,\sigma}\,\|y-y_0\|^2 + \frac{1}{2\,\tau}\,\|x-x_0\|^2 \geq T_K\,\mathcal{L}(\tilde{X}_K;y) - T_K\,\mathcal{L}(x;\tilde{Y}_K) \\ + \frac{1-\tau\,\sigma\,|||A|||^2\,\theta}{2\,\sigma\,\omega^K}\,\|y-y_K\|^2 + \frac{1}{2\,\tau\,\omega^K}\,\|x-x_K\|^2 \\ + \frac{1}{\omega^{K-1}}\,\frac{1}{2\,\tau}\,\|x_K - x_{K-1}\|^2\end{aligned}$$

Proposition 8 (Cas régulier, $\theta < 1$)

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 $|||A|||$. On suppose que g est fortement convexe, de module α et que f est $L_{\nabla f}$-régulière. On pose

$$\forall\,(x,y) \in \mathcal{X}\times\mathcal{Y}, \qquad \mathcal{L}(x;y) = g(x) + \langle y, A\,x\rangle - f^*(y)$$

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

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

où τ, $\sigma > 0$ et $\theta \in \,]\,0\,;1\,[$. Si $\sigma = \tau\,\alpha\,L_{\nabla f}$, $\tau\,\sigma\,|||A|||^2 \leq 1$ et $\theta \geq 1/(1+\tau\,\alpha)$, et ω tel que

$$\frac{1+\theta}{2+\tau\,\alpha} \leq \omega \leq \theta$$

alors la suite $(x_k)_{k\in\mathbb{N}}$ converge vers l'unique solution x^* du problème

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

et il existe $C > 0$ tel que

$$\forall\,k\in\mathbb{N}^*, \qquad \|x^* - x_k\|^2 \leq \omega^k\,C$$

De plus, la suite $((x_k,y_k))_{k\in\mathbb{N}}$ converge vers le point-selle (x^*,y^*) de $\mathcal{L}$ et il existe $C' > 0$ tel que

$$\forall\,k\in\mathbb{N}^*, \qquad \|x^*-x_k\|^2 + \|y^*-y_k\|^2 \leq \omega^k\,C'$$

Remarque : Ainsi, dans le cas régulier, la convergence est linéaire.

Démonstration : On applique le résultat du 7 au point-selle (x^*, y^*) ; par définition du point-selle, on obtient la minoration suivante

$$\frac{1}{2\sigma}\,\|y^*-y_0\|^2 + \frac{1}{2\tau}\,\|x^*-x_0\|^2 \geq \frac{1-\tau\,\sigma\,|||A|||^2\,\omega}{2\,\sigma\,\omega^K}\,\|y^*-y_K\|^2 + \frac{1}{2\,\tau\,\omega^K}\,\|x^*-x_K\|^2$$

Il suffit alors de multiplier par θ^K, puis de faire tendre K vers $+\infty$. ■

Il est possible d'établir des règles de choix pour les pas de temps σ et τ conduisant à des valeurs minimales pour le paramètre θ, c'est-à-dire au meilleur taux de convergence théorique pour l'algorithme de Chambolle–Pock.

4.4 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). Il consiste, dans le cas d'une image en niveaux de gris, à résoudre le problème de minimisation suivant :

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

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

$$g = u + n$$

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 g. 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 :

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

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

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

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

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 g.

Notons J la fonction objectif de ce problème. Cette fonction est convexe. Le terme de fidélité g est simple, tandis que le terme de régularité de la forme $f(\nabla^h u)$ avec f la norme euclidienne. En effet,

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

La règle de Fermat s'écrit alors

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

par ailleurs, en utilisant l'identité de Moreau généralisée, on prouve que $\mathrm{prox}_{\sigma f^*}$ est la projection orthogonale sur la boule unité. Les itérations de l'algorithme de Chambolle–Pock s'écrivent alors

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