

Convergence d'algorithmes primaux-duaux application à l'ADMM

Introduction	165
5.1 La méthode des directions alternées	166
5.1.1 Position du problème	166
5.1.2 L'algorithme ADMM	167
5.1.3 Liens avec d'autres algorithmes	169
5.2 Algorithme PDHG : cas régulier	172
5.2.1 Algorithme PDHG	172
5.2.2 Résultats de convergence dans le cas régulier	173
5.2.3 Choix des paramètres	179
5.2.4 Relaxation sur la variable duale	183
5.3 Application à l'ADMM	184
5.3.1 Lien entre PDHG avec sur-relaxation et ADMM	184
5.3.2 Taux de convergence de l'ADMM classique	185
5.3.3 Variante proposée de l'ADMM	187
5.4 Exemples numériques	191
5.4.1 Algorithme FISTA	191
5.4.2 Comparaison théorique avec FISTA et PDHG	191
5.4.3 Premier test quadratique	191
5.4.4 Application au débruitage TV-Huber	196
5.4.5 Discussion	198
Conclusion	200

Introduction

La méthode des directions alternées (en anglais ADMM pour *Alternating Direction Method of Multipliers*) a été initialement proposée en 1976 par GABAY et MERCIER [9] et en 1975 par GLOWINSKI et MARROCCO [10]. C'est l'une des nombreuses méthodes [7] proposées pour résoudre les problèmes composites de la forme

$$\min_{x \in X} \left\{ G(x) + F(Kx) \right\}.$$

Cette méthode a fait l'objet d'une étude intensive ces dernières années, en particulier à propos de sa convergence. Les résultats de convergence proposés se limitent généralement à des hypothèses restrictives sur le rang de la matrice K ainsi que sur la régularité des fonctions F et G . L'ADMM est par ailleurs reliée à d'autres méthodes classiquement rencontrées en optimisation convexe, comme les méthodes DOUGLAS-RACHFORD ou PEACEMAN-RACHFORD, ainsi qu'on le verra au paragraphe 5.1.3, ce qui a permis en particulier d'en proposer de nombreuses variantes.

Dans ce chapitre, on commencera par proposer une courte revue de la méthode des directions alternées classique, ainsi que ses liens avec d'autres méthodes, et certaines des variantes récemment proposées dans la littérature (section 5.1). On montrera ensuite dans la section 5.2 que l'algorithme classique peut se réécrire comme un algorithme primal-dual proximal. On propose alors à une preuve de la convergence de ce dernier dans le cas régulier, d'où on peut dériver un résultat de convergence sur la méthode des directions alternées. Enfin, la section 5.3 présente une variante simple de l'ADMM dont on démontre qu'elle présente un meilleur taux de convergence théorique.

5.1 La méthode des directions alternées

On présente dans cette section le problème étudié dans ce chapitre et on introduit la méthode des multiplicateurs, dont on rappelle quelques résultats connus. Une courte revue d'algorithmes existants liés à cette méthode est enfin présentée.

5.1.1 Position du problème

Problème composite Nous étudions dans ce chapitre les problèmes composites de la forme

$$\min_{x \in X} \left\{ G(x) + F(Kx) \right\}. \quad (5.1)$$

Pour l'instant, on ne fait aucune hypothèse particulière sur les fonctions convexes et propres $G : X \rightarrow \mathbb{R} \cup \{+\infty\}$ et $F : Z \rightarrow \mathbb{R} \cup \{+\infty\}$, ni sur l'opérateur linéaire continu $K : X \rightarrow Z$ (de norme finie), si ce n'est l'existence d'une solution au problème (5.1). Dans ce qui suit, X et Z désignent des espaces de HILBERT de dimensions finies. Le problème sans contrainte (5.1) peut également s'écrire sous la forme d'un problème sous contrainte linéaire :

$$\min_{\substack{(x,z) \in X \times Z \\ Kx=z}} \left\{ G(x) + F(z) \right\}. \quad (5.2)$$

Cette écriture permet en outre de rendre le lagrangien du problème séparable en x et z , le couplage se faisant dans la contrainte linéaire.

De nombreuses méthodes ont été proposées pour résoudre ce problème. Elles reposent généralement sur une stratégie d'éclatement qui vise à découpler les deux termes de la fonction, puis elles exploitent des hypothèses de régularité sur les fonctions ou de rang sur l'opérateur. On s'intéressera dans cette section à l'ADMM (et ses variantes), tandis qu'une autre méthode sera présentée dans la section suivante.

Lagrangien augmenté Pour résoudre le problème (5.2), la méthode ADM propose de considérer le lagrangien augmenté, défini pour tout $(x,y,z) \in X \times Z \times Z$ par

$$L_\lambda(x,y,z) = G(x) + F(z) + \langle y, Kx - z \rangle + \frac{1}{2\lambda} \|Kx - z\|^2$$

où le paramètre λ est supposé strictement positif. Le lagrangien augmenté possède les mêmes points selles que le lagrangien

$$L(x, y, z) = G(x) + F(z) + \langle y, Kx - z \rangle$$

où y est un multiplicateur de LAGRANGE permettant de tenir compte de la contrainte égalité $Kx = y$. L'ajout du terme quadratique permet d'accélérer la recherche de ces derniers. On peut ainsi s'intéresser au problème primal-dual

$$\min_{(x,z) \in X \times Z} \sup_{y \in Z} L_\lambda(x, y, z) \quad (5.3)$$

dont une solution (x^*, y^*, z^*) fournit une solution x^* au problème (5.1) (avec $Kx^* = z^*$).

5.1.2 L'algorithme ADMM

La méthode des multiplicateurs Pour résoudre le problème de recherche de point-selle (5.3), une première méthode, appelée *méthode des multiplicateurs*, a été proposée. Elle consiste à alterner une minimisation partielle du lagrangien augmenté en les variables primales (x, z) et une étape de montée de gradient en la variable duale y . Cela conduit à l'algorithme suivant : on initialise avec $y_0 \in Z^*$, puis on effectue pour tout $n \in \mathbb{N}$ les mises-à-jours suivantes

$$\begin{cases} (x_{n+1}, z_{n+1}) = \operatorname{argmin}_{(x,z) \in X \times Z} L_\lambda(x, y_n, z) \\ y_{n+1} = y_n + \sigma \frac{\partial L_\lambda}{\partial y}(x_{n+1}, y_n, z_{n+1}) \end{cases}$$

qui s'écrit encore

$$\begin{cases} (x_{n+1}, z_{n+1}) = \operatorname{argmin}_{(x,z) \in X \times Z} \left\{ G(x) + F(z) + \langle y_n, Kx - z \rangle + \frac{1}{2\lambda} \|Kx - z\|^2 \right\} \\ y_{n+1} = y_n + \sigma (Kx_{n+1} - z_{n+1}). \end{cases}$$

On notera que la méthode des multiplicateurs, telle qu'elle est écrite ici, suppose l'existence d'un minimum unique pour la fonction $(x, z) \mapsto L_\lambda(x, y_n, z)$, ce qui n'est *a priori* pas le cas.

Directions alternées La méthode des multiplicateurs transforme le problème de minimisation du problème composite (5.1) en une série de problèmes de minimisations où les variables sont à nouveau couplées, donc plus difficiles à résoudre. Pour s'affranchir de cette difficulté, on peut choisir de remplacer la minimisation partielle du lagrangien en (x, z) en deux problèmes de minimisations séparées en x et z . Les deux minimisations peuvent se faire en parallèle à partir de l'itération (x_n, y_n, z_n) :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} L_\lambda(x, y_n, z_n) \\ z_{n+1} = \operatorname{argmin}_{z \in Z} L_\lambda(x_n, y_n, z) \end{cases}$$

ou bien, dans le cas de l'ADMM, de manière successive, en mettant à jour l'une de deux variables primales entre les deux minimisations partielles, par exemple :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} L_\lambda(x, y_n, z_n) \\ z_{n+1} = \operatorname{argmin}_{z \in Z} L_\lambda(x_{n+1}, y_n, z). \end{cases}$$

L'ADMM nous amène donc à considérer l'algorithme suivant :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} \left\{ G(x) + \langle y_n, Kx \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 \right\} \\ z_{n+1} = \operatorname{argmin}_{z \in Z} \left\{ F(z) - \langle y_n, z \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z\|^2 \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda} (Kx_{n+1} - z_{n+1}). \end{cases}$$

où, pour plus de lisibilité, on a supprimé les termes constants superflus. Ainsi qu'on le montrera à la section suivante, le choix $\sigma = 1/\lambda$ pour le pas de montée est un choix naturel pour cet algorithme. Dans [1], les auteurs montrent par ailleurs que cette valeur apparaît lorsque l'on construit l'algorithme ADMM en écrivant les conditions d'optimalité pour les points selles du lagrangien.

Une fois encore, l'algorithme proposé suppose l'unicité des solutions des deux problèmes de minimisations partielles, ce qui n'est pas assuré pour le premier problème (le second est un problème fortement convexe donc strictement convexe). Il faut pour cela introduire d'autres hypothèses, généralement sur la forte convexité de G ou une hypothèse sur le rang de K , pour assurer la forte convexité du premier problème.

Convergence La plupart des résultats de convergence, dont certains seront présentés plus bas, repose sur des hypothèses plus ou moins restrictives concernant l'opérateur K ou les fonctions G et F . Les cas les plus intéressants sont ceux qui transforment les deux problèmes de minimisations partielles en des problèmes fortement convexes, soit grâce à la forte convexité des fonctions, soit grâce au rang de la matrice K (qui rend le terme quadratique $x \mapsto \|Kx - z_n\|^2$ fortement convexe). Ce sont naturellement les conditions les plus restrictives.

Un premier résultat très général [3] assure que

Théorème 13 *Si le lagrangien L admet un point-selle, alors on a trois résultats de convergence pour la suite des itérées de l'algorithme ADMM*

1. *convergence du résiduel : si on pose $r_n = Kx_n - z_n$ pour tout entier n , la suite $(r_n)_{n \in \mathbb{N}}$ converge vers 0 ;*
2. *convergence vers le minimum : la suite $(G(x_n) + F(z_n))_{n \in \mathbb{N}}$ tend vers le minimum de la fonction $x \mapsto G(x) + F(Kx)$;*
3. *convergence de la variable duale : la suite $(y_n)_{n \in \mathbb{N}}$ converge vers y^* .*

On notera que, sans hypothèse additionnelle, la convergence des variables primales n'est pas assurée.

Il faut par ailleurs distinguer les résultats de convergence dit *ergodiques*, qui s'appliquent à la suite définie comme une moyenne arithmétique (pondérée ou non) des variables générées par l'algorithme, des résultats non ergodiques, qui concernent directement la suite des variables $(x_n, y_n, z_n)_{n \in \mathbb{N}}$.

Généralisation La méthode des directions alternées a été en réalité introduite pour résoudre le problème plus général

$$\min_{\substack{(x,z) \in X \times Z \\ Ax=Bz}} \left\{ G(x) + F(z) \right\}.$$

On se ramène au cas étudié ici lorsque $A = K$ et $B = I_Z$ (l'identité). La raison pour laquelle on se place dans ce chapitre dans ce cas particulier tient au fait que nous ferons plus tard des hypothèses de régularité qui se traduiront dans le cas général par la forte convexité de la fonction $y \mapsto F^*(B^*(y))$. Cette hypothèse se traduit en particulier par l'inversibilité de l'opérateur B , qui permet de se ramener dans le cas particulier en réécrivant la contrainte linéaire $B^{-1}Ax = z$.

5.1.3 Liens avec d'autres algorithmes

On a déjà vu le lien entre l'ADMM et la méthode des multiplicateurs. Dans ce paragraphe, on va expliciter ses liens avec d'autres méthodes fréquemment rencontrées en optimisation convexe, ainsi que quelques unes des variantes proposées dans la littérature.

Méthode Douglas-Rachford Montrons que l'ADMM n'est rien d'autre que l'algorithme de DOUGLAS-RACHFORD appliqué au problème dual. Celui-ci est donné par

$$\max_{y \in Z} \left\{ -G^*(-K^*y) - F^*(y) \right\} \quad (5.4)$$

qui est équivalent au problème

$$\min_{y \in Z} \left\{ g(y) + f(y) \right\}$$

avec $g(y) = G^*(-K^*y)$ et $f(y) = F^*(y)$. La méthode de DOUGLAS-RACHFORD résout ce genre de problème à l'aide de l'algorithme

$$\begin{cases} p_{n+1} &= \text{PROX}_{g/\lambda}(2y_n - q_n) \\ q_{n+1} &= q_n + p_{n+1} - y_n \\ y_{n+1} &= \text{PROX}_{f/\lambda}(q_{n+1}) \end{cases}$$

pour un certain $\lambda > 0$.

Montrons que cet algorithme, appliqué au problème dual (5.4), devient la méthode des directions alternées. Commençons par considérer la mise-à-jour de la variable duale y . L'identité de MOREAU assure que

$$y_{n+1} = q_{n+1} - \frac{1}{\lambda} \text{PROX}_{\lambda f^*}(\lambda q_{n+1}). \quad (5.5)$$

Intéressons-nous ensuite à la mise-à-jour de la première variable auxiliaire p . Par définition de l'opérateur proximal, puis de la conjuguée convexe de la fonction G , on a

$$\begin{aligned} p_{n+1} &= \operatorname{argmin}_{p \in Z} \left\{ \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + G^*(-K^*p) \right\} \\ &= \operatorname{argmin}_{p \in Z} \left\{ \sup_{x \in X} \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + \langle -K^*p, x \rangle - G(x) \right\} \end{aligned}$$

Aussi, le problème que l'on cherche à résoudre est le problème de recherche de point-selle

$$\min_{p \in Z} \sup_{x \in X} \left\{ \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + \langle -K^*p, x \rangle - G(x) \right\}$$

qui est équivalent à

$$\max_{x \in X} \inf_{p \in Z} \left\{ \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + \langle -K^*p, x \rangle - G(x) \right\}$$

Or, pour tout $x \in X$ et $p \in Z$ donnés, on a

$$\begin{aligned} \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + \langle -K^*p, x \rangle &= \frac{\lambda}{2} \left\| 2y_n - q_n + \frac{Kx}{\lambda} - p \right\|^2 \\ &\quad - \frac{\lambda}{2} \left\| 2y_n - q_n + \frac{Kx}{\lambda} \right\|^2 + \frac{\lambda}{2} \|2y_n - q_n\|^2 \end{aligned}$$

Cette quantité est minimale par rapport à p lorsque le premier terme s'annule, ce qui est le cas pour $p_{n+1} = 2y_n - q_n + Kx/\lambda$. On obtient alors, pour tout $x \in X$ donné,

$$\begin{aligned} \inf_{p \in Z} \left\{ \frac{\lambda}{2} \|2y_n - q_n - p\|^2 + \langle -K^*p, x \rangle - G(x) \right\} \\ = -\frac{\lambda}{2} \left\| 2y_n - q_n + \frac{Kx}{\lambda} \right\|^2 + \frac{\lambda}{2} \|2y_n - q_n\|^2 - G(x). \end{aligned}$$

Finalement, la mise-à-jour de la variable auxiliaire p devient

$$p_{n+1} = 2y_n - q_n + \frac{1}{\lambda} Kx_{n+1}$$

où la variable x_{n+1} est donnée par

$$\begin{aligned} x_{n+1} &= \operatorname{argmax}_{x \in X} \left\{ -\frac{\lambda}{2} \left\| 2y_n - q_n + \frac{Kx}{\lambda} \right\|^2 + \frac{\lambda}{2} \|2y_n - q_n\|^2 - G(x) \right\} \\ &= \operatorname{argmin}_{x \in X} \left\{ \frac{\lambda}{2} \left\| 2y_n - q_n + \frac{Kx}{\lambda} \right\|^2 + G(x) \right\}. \end{aligned} \quad (5.6)$$

L'introduction de la nouvelle variable x permet de réécrire la mise-à-jour de la seconde variable auxiliaire q :

$$q_{n+1} = y_n + \frac{1}{\lambda} Kx_{n+1}. \quad (5.7)$$

Revenons à la mise-à-jour de la variable duale y , donnée par (5.5), et que l'on réécrit à l'aide de la nouvelle variable x :

$$y_{n+1} = y_n + \frac{1}{\lambda} \left(Kx_{n+1} - \operatorname{prox}_{\lambda f^*}(\lambda y_n + Kx_{n+1}) \right). \quad (5.8)$$

Posons $z_{n+1} = \operatorname{prox}_{\lambda f^*}(\lambda y_n + Kx_{n+1})$; la définition de l'opérateur proximal assure que

$$\begin{aligned} z_{n+1} &= \operatorname{argmin}_{z \in Z} \left\{ \frac{1}{2\lambda} \|z - (\lambda y_n + Kx_{n+1})\|^2 + F(z) \right\} \\ &= \operatorname{argmin}_{z \in Z} \left\{ \frac{1}{2\lambda} \|z - Kx_{n+1}\|^2 - \langle y_n, z \rangle + F(z) \right\}. \end{aligned} \quad (5.9)$$

L'introduction de cette nouvelle variable primale permet de réécrire (5.8) et (5.7) respectivement :

$$y_{n+1} = y_n + \frac{1}{\lambda} (Kx_{n+1} - z_{n+1}) \quad \text{et} \quad q_{n+1} = y_{n+1} + \frac{1}{\lambda} z_{n+1} \quad (5.10)$$

ce qui assure que la mise-à-jour (5.6) de la variable duale x se lit

$$\begin{aligned} x_{n+1} &= \operatorname{argmin}_{x \in X} \left\{ \frac{\lambda}{2} \left\| y_n + \frac{Kx - z_n}{\lambda} \right\|^2 + G(x) \right\} \\ &= \operatorname{argmin}_{x \in X} \left\{ \frac{1}{2\lambda} \|Kx - z_n\|^2 + \langle Kx, y_n \rangle + G(x) \right\}. \end{aligned} \quad (5.11)$$

Finalement, le calcul de y (donné par la première équation dans (5.10)) ne nécessite que les calculs des deux variables primales x et z , donc les mises-à-jours sont données par (5.11) et (5.9). C'est précisément l'algorithme ADMM.

ADMM proximal (PADMM) Au lieu de calculer exactement les deux minimisations dans l'algorithme ADMM, on peut se contenter d'une minimisation approchée, en ajoutant un terme quadratique dans les deux lagrangiens considérés [1] :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} \left\{ G(x) + \langle y_n, Kx \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 + \frac{1}{2} \|x - x_n\|^2 \right\} \\ z_{n+1} = \operatorname{argmin}_{z \in Z} \left\{ F(z) - \langle y_n, z \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z\|^2 + \frac{1}{2} \|z - z_n\|^2 \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda} (Kx_{n+1} - z_{n+1}). \end{cases}$$

On voit alors que cela revient à introduire deux opérateurs proximaux pour les mises-à-jours des deux variables primales : c'est pourquoi on parle d'*ADMM proximal*. Un intérêt majeur de l'ajout de ces termes quadratiques est de rendre les deux problèmes associés fortement convexes, donc d'assurer l'unicité de la solution.

Utilisation des distances de Bregman On peut remplacer dans l'algorithme PADMM les opérateurs proximaux par des itérations de BREGMAN :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} \left\{ G(x) + \langle y_n, Kx \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 + D_{x_n}(x, x_n) \right\} \\ z_{n+1} = \operatorname{argmin}_{z \in Z} \left\{ F(z) - \langle y_n, z \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z\|^2 + D_{z_n}(z, z_n) \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda} (Kx_{n+1} - z_{n+1}). \end{cases}$$

Les plus courantes consistent à considérer deux matrices positives P et Q pour définir la distance de BREGMAN $\|\cdot\|_P = \langle P\cdot, \cdot \rangle$ (*idem* pour $\|\cdot\|_Q$). On choisit généralement les deux matrices P et Q de la forme $P = I - \alpha {}^tMM$ (voir le chapitre suivant).

ADMM relaxé Dans un paragraphe précédent, on a montré que l'ADMM est en réalité la méthode de DOUGLAS-RACHFORD appliquée au problème dual (5.4). Or, une variante de cet algorithme, appelée PEACEMAN-RACHFORD relaxé, est donnée par

$$\begin{cases} p_{n+1} = \operatorname{prox}_{g/\lambda}(2y_n - q_n) \\ q_{n+1} = q_n + 2\beta_n(p_{n+1} - y_n) \\ y_{n+1} = \operatorname{prox}_{f/\lambda}(q_{n+1}) \end{cases}$$

où la suite $(\beta_n)_n$ est à valeurs dans $]0; 1]$. On retrouve la méthode de DOUGLAS-RACHFORD lorsque cette suite est identiquement égale à $1/2$, tandis que le cas particulier $\beta_n = 1$ est appelé méthode de PEACEMAN-RACHFORD. L'idée est donc d'appliquer

la méthode de PEACEMAN-RACHFORD relaxé au problème dual pour obtenir une version relaxée d'ADMM. On obtient alors l'algorithme suivant :

$$\left\{ \begin{array}{l} x_{n+1} = \operatorname{argmin}_{x \in X} L_\lambda(x, y_n, z_n) \\ y_{n+1/2} = y_n + \frac{2\beta_n - 1}{\lambda} \frac{\partial L_\lambda}{\partial y}(x_{n+1}, y_n, z_n) \\ z_{n+1} = \operatorname{argmin}_{z \in Z} L_\lambda(x_{n+1}, y_{n+1/2}, z) \\ y_{n+1} = y_{n+1/2} + \frac{1}{\lambda} \frac{\partial L_\lambda}{\partial y}(x_{n+1}, y_{n+1/2}, z_{n+1}) \end{array} \right.$$

que l'on peut interpréter comme des minimisations partielles, avec après chaque mise-à-jour d'une variable primale un pas de montée (ou éventuellement de descente lorsque $\beta_n < 1/2$) de gradient (soit une de plus que dans le cas non relaxé, où $\beta_n = 1/2$).

5.2 Algorithme PDHG : cas régulier

On présente dans cette section une nouvelle preuve de convergence de l'algorithme primal-dual PDHG sur-relaxé, dans le cas particulier où les fonctions considérées sont fortement convexes. La preuve est inspirée de celle proposée dans [4], mais où les conditions sur les différents paramètres ont été choisies les moins contraignantes possibles. On obtient ainsi des taux de convergence optimaux meilleurs, qui sont ceux présentés dans [5] avec une preuve différente.

5.2.1 Algorithme PDHG

Soient X et Y deux espaces de HILBERT réels de dimension finie. On considère à nouveau dans cette section le problème composite

$$\min_{x \in X} f(Ax) + g(x) \quad (5.12)$$

où $f : X \rightarrow \mathbb{R} \cup \{+\infty\}$ et $g : Y \rightarrow \mathbb{R} \cup \{+\infty\}$ sont des fonctions propres, convexes et s.c.i.. L'opérateur linéaire $A : X \rightarrow Y$ est supposé continu, d'adjoint A^* . On suppose par ailleurs qu'il est borné, de norme L

$$L = \|A\| = \sup_{x \in X, \|x\| \leq 1} \|Ax\|.$$

En utilisant la conjuguée convexe $f^* : Y \rightarrow \mathbb{R} \cup \{+\infty\}$ de la fonction f , on obtient une nouvelle formulation primale-duale du problème (5.12), donnée par le problème de recherche de point-selle

$$\min_{x \in X} \sup_{y \in Y} g(x) + \langle Ax, y \rangle - f^*(y). \quad (5.13)$$

On pose $\mathcal{L}(x, y) = g(x) + \langle Ax, y \rangle - f^*(y)$ le lagrangien du problème (5.13).

Pour résoudre le problème primal-dual (5.13), on se propose d'étudier l'algorithme suivant :

$$\left\{ \begin{array}{l} y_0 \in Y \\ x_0 \in X \\ \bar{x}_0 = x_0 \end{array} \right. \quad \text{et} \quad \forall n \in \mathbb{N}, \quad \left\{ \begin{array}{l} y_{n+1} = \operatorname{prox}_{\sigma f^*}(y_n + \sigma A\bar{x}_n) \\ x_{n+1} = \operatorname{prox}_{\tau g}(x_n - \tau A^* y_{n+1}) \\ \bar{x}_{n+1} = x_{n+1} + \theta (x_{n+1} - x_n) \end{array} \right. \quad (5.14)$$

où les pas de temps $\sigma > 0$ et $\tau > 0$ et le paramètre de relaxation $0 < \theta \leq 1$ seront spécifiés plus tard. Quand $\theta = 0$, cet algorithme est connu sous le nom PDHG (pour *Primal Dual Hybrid Gradient*) [14]. L'étape de sur-relaxation a été introduite par [13] (pour la minimisation de la fonctionnelle MUMFORD-SHAH), puis étudiée dans un cadre plus large dans [8], puis plus récemment dans [5]. La mise-à-jour de la variable duale y peut être interprétée comme un pas de montée de gradient proximale, tandis que la mise-à-jour de la variable primale est un pas de descente de gradient proximale.

5.2.2 Résultats de convergence dans le cas régulier

Désormais, les fonctions f^* et g sont supposées fortement convexes, de paramètre respectif $\delta > 0$ et $\gamma > 0$. On rappelle que l'opérateur A est de norme L . On définit alors le *conditionnement* de la fonction $g + f(A \cdot)$ comme étant la quantité $\kappa = L^2/(\gamma\delta)^1$. On a le résultat de convergence ergodique suivant :

Théorème 14 *Supposons que le problème (5.13) admet une solution, notée (x^*, y^*) . Si on choisit les paramètres $\tau > 0$, $\sigma > 0$ et $0 < \theta \leq 1$ tels que*

$$\max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\sigma\delta + 1} \right\} \leq \theta \leq \frac{1}{L^2\tau\sigma} \quad (5.15)$$

alors, pour tout ω vérifiant

$$\max \left\{ \frac{1}{\tau\gamma + 1}, \frac{\theta + 1}{\sigma\delta + 2} \right\} \leq \omega \leq \theta \quad (5.16)$$

on a la majoration suivante pour tout $N \in \mathbb{N}$ et pour tout $(x, y) \in X \times Y$:

$$\begin{aligned} 0 \leq & \frac{1}{2\tau} \|x_N - x\|^2 + (1 - \omega L^2\tau\sigma) \frac{1}{2\sigma} \|y_N - y\|^2 \\ & + \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) \\ & \leq \frac{\omega^N}{2\tau} \|x_0 - x\|^2 + \frac{\omega^N}{2\sigma} \|y_0 - y\|^2. \end{aligned}$$

Posons maintenant
$$T_N = \sum_{n=1}^N \frac{1}{\omega^{n-1}} = \frac{1 - \omega^N}{\omega^{N-1}(1 - \omega)}$$

et définissons
$$X_N = \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} x_n \quad \text{et} \quad Y_N = \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} y_n.$$

Alors on a l'encadrement suivant pour tout $(x, y) \in X \times Y$

$$\begin{aligned} 0 \leq & \frac{1 - \omega}{\omega(1 - \omega^N)} \frac{1}{2\tau} \|x - x_N\|^2 + \frac{1 - \omega}{\omega(1 - \omega^N)} (1 - \omega L^2\tau\sigma) \frac{1}{2\sigma} \|y - y_N\|^2 \\ & + \mathcal{L}(X_N, y) - \mathcal{L}(x, Y_N) \\ & \leq \frac{1}{T_N} \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{T_N} \frac{1}{2\sigma} \|y - y_0\|^2. \end{aligned}$$

1. Cette définition recouvre celle usuellement utilisée pour les fonctions fortement convexes et de gradient lipschitziens, qui est alors définie comme étant le rapport de la constante de LIPSCHITZ sur le paramètre de forte convexité. Dans ce cas, le conditionnement est nécessaire supérieur ou égal à 1. Dans le cas plus général considéré ici, le conditionnement peut être inférieur à 1.

DÉMONSTRATION : Commençons par considérer une itération de l'algorithme dans sa forme la plus générale : pour tous $(\bar{y}, \tilde{y}) \in Y^2$ et $(\bar{x}, \tilde{x}) \in X^2$, on pose

$$\begin{cases} \hat{y} = \text{prox}_{\sigma f^*}(\bar{y} + \sigma A\tilde{x}) \\ \hat{x} = \text{prox}_{\tau g}(\bar{x} - \tau A^*\tilde{y}). \end{cases}$$

Optimalité Par définition de l'opérateur proximal, le point \hat{x} est le minimiseur d'une fonction fortement convexe :

$$\hat{x} = \underset{x \in X}{\text{argmin}} \left\{ \frac{1}{2\tau} \|\bar{x} - \tau A^*\tilde{y} - x\|^2 + g(x) \right\}.$$

Les conditions nécessaires d'optimalité du premier ordre assurent que

$$-\frac{1}{\tau} (\hat{x} - \bar{x}) - A^*\tilde{y} \in \partial g(\hat{x}).$$

On obtient de manière similaire (en considérant la définition du point \hat{y})

$$-\frac{1}{\sigma} (\hat{y} - \bar{y}) + A\tilde{x} \in \partial f^*(\hat{y}).$$

Forte convexité Exploitions à présent les hypothèses de régularité. En utilisant la définition de la forte convexité, on obtient (en développant les produits scalaires)

$$g(x) + \frac{1}{2\tau} \|x - \bar{x}\|^2 \geq g(\hat{x}) + \langle A(\hat{x} - x), \tilde{y} \rangle + \frac{1}{2\tau} \|\hat{x} - \bar{x}\|^2 + \frac{1}{2\tau} \|x - \hat{x}\|^2 + \frac{\gamma}{2} \|x - \hat{x}\|^2 \quad (5.17)$$

et

$$f^*(y) + \frac{1}{2\sigma} \|y - \bar{y}\|^2 \geq f^*(\hat{y}) - \langle A\tilde{x}, \hat{y} - y \rangle + \frac{1}{2\sigma} \|\hat{y} - \bar{y}\|^2 + \frac{1}{2\sigma} \|y - \hat{y}\|^2 + \frac{\delta}{2} \|y - \hat{y}\|^2. \quad (5.18)$$

Sommons les deux inégalités (5.17) et (5.18). On obtient après un réarrangement des termes

$$\begin{aligned} \mathcal{L}(\hat{x}, y) - \mathcal{L}(x, \hat{y}) &\leq \frac{1}{2\tau} \|x - \bar{x}\|^2 - \frac{1 + \tau\gamma}{2\tau} \|x - \hat{x}\|^2 - \frac{1}{2\tau} \|\bar{x} - \hat{x}\|^2 \\ &\quad + \frac{1}{2\sigma} \|y - \bar{y}\|^2 - \frac{1 + \sigma\delta}{2\sigma} \|y - \hat{y}\|^2 - \frac{1}{2\sigma} \|\bar{y} - \hat{y}\|^2 \\ &\quad + \langle A(\hat{x} - x), \hat{y} - \tilde{y} \rangle - \langle A(\hat{x} - \tilde{x}), \hat{y} - y \rangle. \end{aligned}$$

Choix de \hat{x} , \bar{x} , \tilde{x} et \hat{y} , \bar{y} , \tilde{y} On spécifie maintenant les différents points impliqués dans l'itération. On choisit $\hat{x} = x_{n+1}$, $\bar{x} = x_n$, $\tilde{x} = x_n + \theta(x_n - x_{n-1})$ pour $1 \geq \theta > 0$ (pour l'instant laissé libre), et $\hat{y} = y_{n+1}$, $\bar{y} = y_n$ et $\tilde{y} = y_{n+1}$. Après simplification, on obtient

$$\begin{aligned} \mathcal{L}(x_{n+1}, y) - \mathcal{L}(x, y_{n+1}) &\leq \frac{1}{2\tau} \|x - x_n\|^2 + \frac{1}{2\sigma} \|y - y_n\|^2 \\ &\quad - \frac{1 + \tau\gamma}{2\tau} \|x - x_{n+1}\|^2 - \frac{1 + \sigma\delta}{2\sigma} \|y - y_{n+1}\|^2 \\ &\quad - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 - \frac{1}{2\sigma} \|y_n - y_{n+1}\|^2 \\ &\quad + \theta \langle A(x_{n-1} - x_n), y - y_{n+1} \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle. \end{aligned} \quad (5.19)$$

Posons $\tau\gamma = \mu > 0$ et $\sigma\delta = \mu' > 0$. Pour tout $n \in \mathbb{N}$, on définit

$$\Delta_n = \frac{1}{2\tau} \|x - x_n\|^2 + \frac{1}{2\sigma} \|y - y_n\|^2.$$

On peut alors réécrire l'inégalité (5.19) avec Δ_n , μ et μ' , ce qui donne

$$\begin{aligned} \mathcal{L}(x_{n+1}, y) - \mathcal{L}(x, y_{n+1}) &\leq \Delta_n - (1 + \mu) \Delta_{n+1} - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 - \frac{1}{2\sigma} \|y_n - y_{n+1}\|^2 \\ &\quad + \theta \langle A(x_{n-1} - x_n), y - y_{n+1} \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle \\ &\quad + \frac{\mu - \mu'}{2\sigma} \|y - y_{n+1}\|^2. \end{aligned} \tag{5.20}$$

Produits scalaires dans (5.20) Majorons les produits scalaires dans l'inégalité (5.20).

On commence par écrire pour tout $0 < \omega \leq \theta$ la décomposition

$$\begin{aligned} \theta \langle A(x_{n-1} - x_n), y - y_{n+1} \rangle &= \omega \langle A(x_{n-1} - x_n), y - y_n \rangle \\ &\quad + \omega \langle A(x_{n-1} - x_n), y_n - y_{n+1} \rangle \\ &\quad + (\theta - \omega) \langle A(x_{n-1} - x_n), y - y_{n+1} \rangle. \end{aligned}$$

Bornons chacun des deux derniers termes. En utilisant l'inégalité de CAUCHY-SCHWARZ puis la définition de la norme L , on a pour tout $\alpha > 0$ (puisque $\omega \geq 0$)

$$\begin{aligned} \omega \langle A(x_{n-1} - x_n), y_n - y_{n+1} \rangle &\leq \omega L \|x_{n-1} - x_n\| \cdot \|y_n - y_{n+1}\| \\ &\leq \omega L \left(\frac{\alpha}{2} \|x_{n-1} - x_n\|^2 + \frac{1}{2\alpha} \|y_n - y_{n+1}\|^2 \right). \end{aligned} \tag{5.21}$$

De même, puisque $\theta - \omega \geq 0$, on a pour tout $\alpha' > 0$

$$(\theta - \omega) \langle A(x_{n-1} - x_n), y - y_{n+1} \rangle \leq (\theta - \omega) L \left(\frac{\alpha'}{2} \|x_{n-1} - x_n\|^2 + \frac{1}{2\alpha'} \|y - y_{n+1}\|^2 \right). \tag{5.22}$$

En injectant dans (5.20) les deux bornes (5.21) et (5.22), on obtient après simplification

$$\begin{aligned} \mathcal{L}(x_{n+1}, y) - \mathcal{L}(x, y_{n+1}) &\leq \Delta_n - (1 + \mu) \Delta_{n+1} \\ &\quad + \left(\omega \frac{\alpha}{2} + (\theta - \omega) \frac{\alpha'}{2} \right) L \|x_{n-1} - x_n\|^2 - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 \\ &\quad + \left(\frac{\omega L}{2\alpha} - \frac{1}{2\sigma} \right) \|y_n - y_{n+1}\|^2 \\ &\quad + \omega \langle A(x_{n-1} - x_n), y - y_n \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle \\ &\quad + \left(\frac{(\theta - \omega) L}{2\alpha'} + \frac{\mu - \mu'}{2\sigma} \right) \|y - y_{n+1}\|^2. \end{aligned}$$

Choisissons $\alpha = \alpha' = \omega L\sigma$. De cette manière, on a $\omega L/\alpha = 1/\sigma$, de sorte que le terme en $\|y_n - y_{n+1}\|^2$ s'annule, ce qui entraîne :

$$\begin{aligned} \mathcal{L}(x_{n+1}, y) - \mathcal{L}(x, y_{n+1}) &\leq \Delta_n - (1 + \mu) \Delta_{n+1} \\ &\quad + \omega \frac{\theta L^2 \tau \sigma}{2\tau} \|x_{n-1} - x_n\|^2 - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 \\ &\quad + \omega \langle A(x_{n-1} - x_n), y - y_n \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle \\ &\quad + \left(\frac{\theta - \omega}{\omega} + \mu - \mu' \right) \frac{1}{2\sigma} \|y - y_{n+1}\|^2. \end{aligned} \quad (5.23)$$

Puisque $1 + \mu = 1/\omega + 1 + \mu - 1/\omega$, on a

$$-(1 + \mu) \Delta_{n+1} = -\frac{1}{\omega} \Delta_{n+1} + \left(\frac{1}{\omega} - \mu - 1 \right) \left(\frac{1}{2\tau} \|x - x_{n+1}\|^2 + \frac{1}{2\sigma} \|y - y_{n+1}\|^2 \right)$$

ce qui implique que le membre de gauche dans (5.23) devient

$$\begin{aligned} \Delta_n - \frac{1}{\omega} \Delta_{n+1} + \omega \frac{\theta L^2 \tau \sigma}{2\tau} \|x_n - x_{n-1}\|^2 - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 \\ + \omega \langle A(x_{n-1} - x_n), y - y_n \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle \\ + \left(\frac{1}{\omega} - \mu - 1 \right) \frac{1}{2\tau} \|x - x_{n+1}\|^2 + \left(\frac{\theta - \omega}{\omega} + \frac{1}{\omega} - \mu' - 1 \right) \frac{1}{2\sigma} \|y - y_{n+1}\|^2. \end{aligned} \quad (5.24)$$

Conditions sur ω , θ , τ et σ L'idée est à présent d'imposer des contraintes sur les valeurs des paramètres de sorte de simplifier l'inégalité précédente en la majorant par une expression plus simple. On commence par imposer que θ , τ et σ vérifient $\theta L^2 \tau \sigma \leq 1$. Ensuite, on choisit θ de sorte que $1/\omega - \mu - 1$ et $(\theta - \omega)/\omega + 1/\omega - \mu' - 1$ soient tous deux négatifs, ce qui entraîne que

$$\frac{1}{\mu + 1} \leq \omega \quad \text{et} \quad \frac{\theta + 1}{\mu' + 2} \leq \omega.$$

Ainsi, on peut majorer le terme de droite dans (5.24) par

$$\begin{aligned} \Delta_n - \frac{1}{\omega} \Delta_{n+1} + \omega \frac{1}{2\tau} \|x_{n-1} - x_n\|^2 - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 \\ + \omega \langle A(x_{n-1} - x_n), y - y_n \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle. \end{aligned}$$

Finalement, si on revient à l'inégalité (5.23), on obtient

$$\begin{aligned} \mathcal{L}(x_n, y) - \mathcal{L}(x, y_n) &\leq \Delta_n - \frac{1}{\omega} \Delta_{n+1} \\ &\quad + \omega \frac{1}{2\tau} \|x_{n-1} - x_n\|^2 - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 \\ &\quad + \omega \langle A(x_{n-1} - x_n), y - y_n \rangle - \langle A(x_n - x_{n+1}), y - y_{n+1} \rangle. \end{aligned} \quad (5.25)$$

Convergence ergodique Multiplions (5.25) par $1/\omega^n$ puis sommons pour n entre 0 et $N - 1$. Si on pose $x^{-1} = x^0$, certains se télescopent, et on obtient alors

$$\begin{aligned} \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) &\leq \Delta_0 - \frac{1}{\omega^N} \Delta_N - \frac{1}{2\tau\omega^{N-1}} \|x_{N-1} - x_N\|^2 \\ &\quad - \frac{1}{\omega^{N-1}} \langle A(x_{N-1} - x_N), y - y_N \rangle. \end{aligned} \quad (5.26)$$

On majore à nouveau le produit scalaire à l'aide de $\beta > 0$,

$$-\frac{1}{\omega^{N-1}} \langle A(x_{N-1} - x_N), y - y_N \rangle \leq \frac{L}{\omega^{N-1}} \left(\frac{\beta}{2} \|x_{N-1} - x_N\|^2 + \frac{1}{2\beta} \|y - y_N\|^2 \right).$$

L'inégalité (5.26) devient alors

$$\begin{aligned} \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) &\leq \Delta_0 - \frac{1}{\omega^N} \Delta_N + \left(\frac{L\beta}{2\omega^{N-1}} - \frac{1}{2\tau\omega^{N-1}} \right) \|x_{N-1} - x_N\|^2 \\ &\quad + \frac{L}{\omega^{N-1}} \frac{1}{2\beta} \|y_N - y^*\|^2. \end{aligned}$$

Si on choisit $\beta = 1/(L\tau)$, le terme en $\|x_{N-1} - x_N\|^2$ disparaît, et on obtient l'inégalité suivante

$$\sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) \leq \Delta_0 - \frac{1}{\omega^N} \Delta_N + \frac{L^2\tau\sigma}{\omega^{N-1}} \frac{1}{2\sigma} \|y - y_N\|^2.$$

Maintenant, on remplace Δ_0 et Δ_n par leurs définitions respectives :

$$\begin{aligned} \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) &\leq \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{2\sigma} \|y - y_0\|^2 \\ &\quad - \frac{1}{\omega^N} \frac{1}{2\tau} \|x - x_N\|^2 - \frac{1}{\omega^N} (1 - \omega L^2\tau\sigma) \frac{1}{2\sigma} \|y - y_N\|^2. \end{aligned}$$

Puisque $\omega L^2\tau\sigma \leq \theta L^2\tau\sigma \leq 1$ et $\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n) \geq 0$ pour tout $n \in \mathbb{N}$, on obtient après un réarrangement des termes

$$\begin{aligned} 0 &\leq \frac{1}{\omega^N} \frac{1}{2\tau} \|x - x_N\|^2 + \frac{1}{\omega^N} (1 - \omega L^2\tau\sigma) \frac{1}{2\sigma} \|y - y_N\|^2 \\ &\quad + \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) \\ &\leq \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{2\sigma} \|y - y_0\|^2. \end{aligned}$$

Divisons maintenant par $T_N \neq 0$, ce qui assure que

$$\begin{aligned} 0 &\leq \frac{1 - \omega}{\omega(1 - \omega^N)} \frac{1}{2\tau} \|x - x_N\|^2 + \frac{1 - \omega}{\omega(1 - \omega^N)} (1 - \omega L^2\tau\sigma) \frac{1}{2\sigma} \|y - y_N\|^2 \\ &\quad + \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) \\ &\leq \frac{1}{T_N} \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{T_N} \frac{1}{2\sigma} \|y - y_0\|^2. \end{aligned} \quad (5.27)$$

Utilisons la convexité du lagrangien \mathcal{L} en sa première variable, et sa concavité en sa seconde variable, pour écrire :

$$0 \leq \mathcal{L}(X_N, y) - \mathcal{L}(x, Y_N) \leq \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)).$$

Ainsi, l'encadrement (5.27) devient

$$\begin{aligned} 0 &\leq \frac{1-\omega}{\omega(1-\omega^N)} \frac{1}{2\tau} \|x - x_N\|^2 + \frac{1-\omega}{\omega(1-\omega^N)} (1 - \omega L^2 \tau \sigma) \frac{1}{2\sigma} \|y - y_N\|^2 \\ &\quad + \mathcal{L}(X_N, y) - \mathcal{L}(x, Y_N) \\ &\leq \frac{1}{T_N} \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{T_N} \frac{1}{2\sigma} \|y - y_0\|^2 \end{aligned} \quad (5.28)$$

ce qui achève la démonstration. ■

REMARQUE : L'encadrement (5.28) permet de démontrer la convergence linéaire des variables x_n et, si $1 \neq \omega L^2 \tau \sigma$, celle des y_n , de taux ω . Mieux encore, écrivons la définition

$$\mathcal{L}(x_{n+1}, y^*) - \mathcal{L}(x^*; y_{n+1}) = g(x_{n+1}) - g(x^*) + f^*(y_{n+1}) - f^*(y^*) + \langle Ax_{n+1}, y^* \rangle - \langle Ax^*, y_{n+1} \rangle.$$

La forte convexité de g et de f^* et les conditions d'optimalité pour x^* et y^* impliquent

$$g(x_{n+1}) - g(x^*) \geq \langle -A^* y^*, x_{n+1} - x^* \rangle + \frac{\gamma}{2} \|x_{n+1} - x^*\|^2$$

et

$$f^*(y_{n+1}) - f^*(y^*) \geq \langle Ax^*, y_{n+1} - y^* \rangle + \frac{\delta}{2} \|y_{n+1} - y^*\|^2.$$

Ainsi, on obtient

$$\frac{\gamma}{2} \|x_{n+1} - x^*\|^2 + \frac{\delta}{2} \|y_{n+1} - y^*\|^2 \leq \mathcal{L}(x_{n+1}, y^*) - \mathcal{L}(x^*, y_{n+1})$$

puisque la somme des produits scalaires s'annule. Ainsi, si on choisit d'abandonner le contrôle sur le *prima-dual gap*, l'inégalité (5.19) devient

$$\begin{aligned} 0 &\leq \frac{1}{2\tau} \|x^* - x_n\|^2 + \frac{1}{2\sigma} \|y^* - y_n\|^2 \\ &\quad - \frac{1+2\tau\gamma}{2\tau} \|x^* - x_{n+1}\|^2 - \frac{1+2\sigma\delta}{2\sigma} \|y^* - y_{n+1}\|^2 \\ &\quad - \frac{1}{2\tau} \|x_n - x_{n+1}\|^2 - \frac{1}{2\sigma} \|y_n - y_{n+1}\|^2 \\ &\quad + \theta \langle A(x_{n-1} - x_n), y^* - y_{n+1} \rangle - \langle A(x_n - x_{n+1}), y^* - y_{n+1} \rangle. \end{aligned} \quad (5.29)$$

Il s'ensuit que tous les calculs depuis (5.19) jusqu'à (5.28) restent valables, avec cette fois μ et μ' remplacés par $\tilde{\mu} = 2\mu$ et $\tilde{\mu}' = 2\mu'$ et sans les termes en \mathcal{L} , de même que dans les conditions sur les paramètres. En d'autres termes, des calculs similaires prouvent le corollaire suivant :

Corollaire 1 *Supposons que le problème (5.13) admet une solution, notée (x^*, y^*) . Si on choisit les paramètres $\tau > 0$, $\sigma > 0$, $0 < \theta \leq 1$ tels que*

$$\max \left\{ \frac{1}{2\tau\gamma + 1}, \frac{1}{2\sigma\delta + 1} \right\} \leq \theta \leq \frac{1}{L^2\tau\sigma} \quad (5.30)$$

alors, pour tout $\tilde{\omega}$ vérifiant

$$\max \left\{ \frac{1}{2\tau\gamma + 1}, \frac{\theta + 1}{2\sigma\delta + 2} \right\} \leq \tilde{\omega} \leq \theta \quad (5.31)$$

on a la majoration suivante pour tout $N \in \mathbb{N}$

$$\|x^* - x_N\|^2 \leq \tilde{\omega}^N \left(\|x^* - x_0\|^2 + \frac{\tau}{\sigma} \|y^* - y_0\|^2 \right). \quad (5.32)$$

De plus, si $\tilde{\omega}L^2\tau\sigma \neq 1$, alors on a également

$$\|y^* - y_N\|^2 \leq \frac{\tilde{\omega}^N}{1 - \tilde{\omega}L^2\tau\sigma} \left(\frac{\sigma}{\tau} \|x^* - x_0\|^2 + \|y^* - y_0\|^2 \right). \quad (5.33)$$

Pour tous τ , σ et θ , les bornes inférieures $1/(2\tau\gamma + 1)$ et $(\theta + 1)/(2\sigma\delta + 2)$ pour $\tilde{\omega}$ sont plus petites que celles obtenues pour ω . Ainsi, le nouveau taux $\tilde{\omega}$, qui n'est *a priori* valable que pour la convergence des itérées, est meilleur le taux global ω (qui lui est également valable pour la convergence du *gap*). Notons par ailleurs que tout choix de paramètres satisfaisant (5.15) est compatible avec les contraintes (5.30), ce qui signifie qu'un choix de paramètres (τ, σ, θ) donné par le théorème 14 assure à la fois la convergence linéaire du *gap* au taux ω donné par (5.16) et celle des itérées au taux $\tilde{\omega}$ donné par (5.31).

5.2.3 Choix des paramètres

Le théorème 14 est valable tant qu'on parvient à choisir les pas de temps τ et σ et le paramètre de relaxation θ qui satisfont les contraintes imposées. La valeur minimale des ω correspondant donne alors le taux de convergence de l'algorithme. On va étudier ici plusieurs choix possibles, ainsi que les taux de convergence associés. On procédera de la manière suivante :

1. On fixe $\tau > 0$.
2. On cherche les conditions sur σ pour que les inégalités (5.15) existent pour au moins un θ (*i.e.* pour que le membre de gauche soit inférieur au membre de droite).
3. On minimise la quantité $(\theta + 1)/(\sigma\delta + 2)$ par rapport à θ vérifiant (5.15) et par rapport à σ vérifiant les conditions déterminées à l'étape précédente.
4. On compare ce minimum à $1/(\tau\gamma + 1)$ puis on en déduit une borne inférieure $\omega^*(\tau)$ pour ω grâce à (5.16).
5. On minimise enfin $\omega^*(\tau)$ par rapport à τ puis on en déduit le taux optimal ω^* ainsi que le taux $\tilde{\omega}$ correspondant sur la convergence des itérées.

On adaptera évidemment cette procédure lorsqu'un ou plusieurs paramètres seront fixés au préalable.

Cas $\theta = 1$ Fixons $\tau > 0$. En remplaçant $\theta = 1$ dans (5.15), on obtient la contrainte suivante sur le pas σ

$$1 \leq \frac{1}{L^2\tau\sigma} \quad (5.34)$$

ce qui implique que $\sigma \leq 1/(L^2\tau)$. La valeur de θ étant fixée, on cherche maintenant à minimiser $1/(\sigma\delta/2 + 1)$ par rapport à σ vérifiant les conditions (5.34). Puisque la fonction $\sigma \mapsto 1/(\sigma\delta/2 + 1)$ est décroissante sur \mathbb{R}^+ , son minimum est atteint lorsque σ est maximal, et vaut

$$\min_{\sigma \text{ vérifiant (5.34)}} \left\{ \frac{1}{\sigma\delta/2 + 1} \right\} = \frac{1}{\delta/(2L^2\tau) + 1}.$$

Comparons cette valeur à $1/(\tau\gamma + 1)$. Il est clair que $1/(\delta/(2L^2\tau) + 1)$ est supérieur à $1/(\tau\gamma + 1)$ dès que $\tau \geq \sqrt{\delta/(2\gamma L^2)}$. Ainsi, la borne inférieure $\omega^*(\tau)$, qui vaut le maximum entre ces deux quantités, est donnée par

$$\omega^*(\tau) = \max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\delta/(2L^2\tau) + 1} \right\} = \begin{cases} \frac{1}{\tau\gamma + 1} & \text{if } 0 < \tau < \sqrt{\delta/(2\gamma L^2)} \\ \frac{1}{\delta/(2L^2\tau) + 1} & \text{if } \tau \geq \sqrt{\delta/(2\gamma L^2)}. \end{cases}$$

Celle-ci est minimale pour $\tau^* = \sqrt{\delta/(2\gamma L^2)}$ et conduit au taux de convergence optimal

$$\omega^* = \omega^*(\tau^*) = \frac{1}{\sqrt{(\gamma\delta)/(2L^2)} + 1} = \frac{1}{\sqrt{1/(2\kappa)} + 1}.$$

Ce taux est atteint pour

$$\tau = \tau^* = \sqrt{\frac{\delta}{2\gamma L^2}} \quad \text{et} \quad \sigma = \frac{1}{L^2\tau^*} = \sqrt{\frac{2\gamma}{\delta L^2}}.$$

On vérifie par ailleurs que le taux de convergence sur les itérées $\tilde{\omega}$ associé à ce choix de paramètres vérifie dans ce cas $\tilde{\omega} = \omega^*$ et que cette valeur est minimale étant donnée la contrainte $\theta = 1$.

Meilleur taux de convergence ($\theta < 1$) On cherche à présent à déterminer les valeurs des paramètres conduisant au taux de convergence minimal.

Théorème 15 *Le meilleur taux de convergence dans le théorème 14 est obtenu en choisissant*

$$\tau = \frac{\delta}{2L^2} \left(1 + \sqrt{1 + \frac{4L^2}{\gamma\delta}} \right) \quad \text{et} \quad \sigma = \frac{\gamma}{2L^2} \left(1 + \sqrt{1 + \frac{4L^2}{\gamma\delta}} \right)$$

pour les pas de temps, ainsi que, pour le paramètre de relaxation,

$$\theta = \frac{\sqrt{1 + \frac{4L^2}{\gamma\delta}} - 1}{\sqrt{1 + \frac{4L^2}{\gamma\delta}} + 1} = \frac{\sqrt{1 + 4\kappa} - 1}{\sqrt{1 + 4\kappa} + 1} < 1.$$

Ces valeurs satisfont $\tau\gamma = \sigma\delta$ et $\theta L^2\tau\sigma = 1$. On a par ailleurs $\omega^ = \theta$.*

DÉMONSTRATION : Fixons $\tau > 0$ et déterminons les conditions que σ doit vérifier pour assurer l'existence de θ satisfaisant (5.15). Il existe de telles valeurs de θ si

$$\frac{1}{\tau\gamma + 1} \leq \frac{1}{L^2\tau\sigma} \quad \text{et} \quad \frac{1}{\sigma\delta + 1} \leq \frac{1}{L^2\tau\sigma}.$$

qui s'écrit également

$$\sigma \leq \frac{1}{L^2\tau} + \frac{\gamma}{L^2} \quad \text{et} \quad (L^2\tau - \delta)\sigma \leq 1.$$

Déterminons ensuite les conditions sur σ pour que ces inégalités existent. Si $L^2\tau - \delta$ est négatif, *i.e.* $\tau \leq \delta/L^2$, la seconde inégalité est toujours vérifiée. La première inégalité fournit donc une borne supérieure pour σ . Ainsi, étudions le cas $L^2\tau - \delta$ est strictement positif, *i.e.* $\tau > \delta/L^2$. Le pas σ doit alors vérifier les deux majorations

$$\sigma \leq \frac{1}{L^2\tau} + \frac{\gamma}{L^2} \quad \text{et} \quad \sigma \leq \frac{1}{L^2\tau - \delta}. \quad (5.35)$$

Comparons ces deux majorants. Puisque

$$\frac{1}{L^2\tau} + \frac{\gamma}{L^2} - \frac{1}{L^2\tau - \delta} = \frac{\gamma L^2\tau^2 - \gamma\delta\tau - \delta}{L^2\tau(L^2\tau - \delta)}$$

avec $L^2\tau(L^2\tau - \delta)$ positif, $1/(L^2\tau) + \gamma/L^2$ est supérieur à $1/(L^2\tau - \delta)$ si et seulement si $\gamma L^2\tau^2 - \gamma\delta\tau - \delta \geq 0$, c'est-à-dire si et seulement si $\tau \geq \tau^*$, donné par

$$\tau^* = \frac{\delta}{2L^2} \left(1 + \sqrt{1 + \frac{4L^2}{\gamma\delta}} \right) > \frac{\delta}{L^2}.$$

Ainsi, pour tout $\delta/L^2 < \tau \leq \tau^*$, les conditions (5.35) deviennent $\sigma \leq 1/(L^2\tau) + \gamma/L^2$. Si $\tau > \tau^*$, alors les conditions (5.35) se lisent $\sigma \leq 1/(L^2\tau - \delta)$. En conclusion, le paramètre σ doit vérifier

$$\sigma \leq \begin{cases} \frac{1}{L^2\tau} + \frac{\gamma}{L^2} & \text{si } 0 < \tau \leq \tau^* \\ \frac{1}{L^2\tau - \delta} & \text{si } \tau^* < \tau. \end{cases} \quad (5.36)$$

Maintenant, fixons σ vérifiant (5.36) et minimisons la quantité $(\theta + 1)/(\sigma\delta + 2)$ avec θ vérifiant les contraintes (5.15). La fonction $\theta \mapsto (\theta + 1)/(\sigma\delta + 2)$ atteint son minimum quand θ est minimal. Or, la borne inférieure de θ est donnée par

$$\max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\sigma\delta + 1} \right\}.$$

Calculons explicitement cette quantité. Remarquons tout d'abord que, si $\tau > \delta/L^2$, alors

$$\frac{\delta}{L^2\tau - \delta} \leq \tau\gamma \quad \iff \quad \gamma L^2\tau^2 - \gamma\delta\tau - \delta \geq 0 \quad \iff \quad \tau \geq \tau^*. \quad (5.37)$$

Supposons que $\tau > \tau^*$, ce qui entraîne $\tau > \delta/L^2$. Puisque σ est majoré par $1/(L^2\tau - \delta)$, on en déduit que $\sigma\delta \leq \tau\gamma$, ce qui implique la borne inférieure de θ est donnée dans ce cas par

$$\theta^*(\sigma) = \max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\sigma\delta + 1} \right\} = \frac{1}{\sigma\delta + 1} \quad \text{si} \quad 0 < \sigma \leq \frac{1}{L^2\tau - \delta}. \quad (5.38)$$

Considérons à présent le cas $\tau \leq \tau^*$. Puisque

$$\frac{1}{L^2\tau} + \frac{\gamma}{L^2} \geq \frac{\tau\gamma}{\delta} \iff \gamma L^2\tau^2 - \gamma\delta\tau - \delta \leq 0 \iff \tau \leq \tau^*,$$

on en déduit que la borne inférieure de θ vaut

$$\theta^*(\sigma) = \max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\sigma\delta + 1} \right\} = \begin{cases} \frac{1}{\sigma\delta + 1} & \text{si } 0 < \sigma \leq \frac{\tau\gamma}{\delta} \\ \frac{1}{\tau\gamma + 1} & \text{si } \frac{\tau\gamma}{\delta} < \sigma \leq \frac{1}{L^2\tau} + \frac{\gamma}{L^2}. \end{cases} \quad (5.39)$$

Minimisons $(\theta + 1)/(\sigma\delta + 2)$ par rapport à σ , quand θ est égal à sa borne inférieure $\theta^*(\sigma)$, donnée par (5.38) lorsque $\tau > \tau^*$ et (5.39) sinon. Cela revient à minimiser la quantité suivante par rapport à σ :

$$\frac{\theta^*(\sigma) + 1}{\sigma\delta + 2} = \begin{cases} \frac{1}{\sigma\delta + 1} & \text{si } \tau > \tau^* \text{ ou } \left(\tau \leq \tau^* \text{ et } 0 < \sigma \leq \frac{\tau\gamma}{\delta} \right) \\ \frac{1}{\tau\gamma + 1} \times \frac{\tau\gamma + 2}{\sigma\delta + 2} & \text{si } \left(\tau \leq \tau^* \text{ et } \frac{\tau\gamma}{\delta} < \sigma \leq \frac{1}{L^2\tau} + \frac{\gamma}{L^2} \right). \end{cases}$$

Dans les deux cas, le minimum est atteint quand σ est maximal, égal à sa borne supérieure donnée par (5.36). Ainsi,

$$\min_{\substack{\sigma \text{ vérifiant (5.36)} \\ \theta \text{ vérifiant (5.15)}}} \left\{ \frac{\theta + 2}{\sigma + 1} \right\} = \begin{cases} 1 - \frac{\delta}{L^2\tau} & \text{si } \tau > \tau^* \\ \min \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\tau\gamma + 1} \times \frac{\tau\gamma + 2}{\delta/(L^2\tau) + \delta\gamma/L^2 + 2} \right\} & \text{si } \tau \leq \tau^*. \end{cases}$$

Comparons cette valeur à $1/(\tau\gamma + 1)$, pour en déduire la borne inférieure de $\omega^*(\tau)$:

$$\omega^*(\tau) = \max \left\{ \frac{1}{\tau\gamma + 1}, \min_{\substack{\sigma \text{ vérifiant (5.36)} \\ \theta \text{ vérifiant (5.15)}}} \left\{ \frac{\theta + 2}{\sigma + 1} \right\} \right\}.$$

Grâce à (5.37), on obtient que

$$\omega^*(\tau) = \begin{cases} 1 - \frac{\delta}{L^2\tau} & \text{si } \tau > \tau^* \\ \frac{1}{\tau\gamma + 1} & \text{si } \tau \leq \tau^* \end{cases}$$

qui est minimal pour $\tau = \tau^*$ ce qui conduit au taux optimal :

$$\omega^* = 1 - \frac{\delta}{L^2\tau^*} = \frac{\sqrt{1 + (4L^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4L^2)/(\gamma\delta)} + 1}$$

obtenu pour $\tau = \tau^*$ et $\sigma = \tau^*\gamma/\delta$. On peut par ailleurs vérifier que le taux $\tilde{\omega}$ valable pour la convergence de la variable primale est dans ce cas donné par

$$\tilde{\omega} = \frac{1}{\tau^*\gamma/2 + 1} = \frac{\sqrt{1 + (4L^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4L^2)/(\gamma\delta)} + 3} = \frac{\sqrt{1 + 4\kappa} - 1}{\sqrt{1 + 4\kappa} + 3}. \blacksquare$$

5.2.4 Relaxation sur la variable duale

La démonstration précédente peut être adaptée lorsque l'on inverse les rôles des variables primale et duale, c'est-à-dire lorsque l'on considère le problème équivalent

$$\min_{y \in Y} \sup_{x \in X} \left\{ F^*(y) + \langle (-K^*)y, x \rangle - G(x) \right\}.$$

Dans ce cas, la variable primale est mise-à-jour avant la variable duale, et la relaxation est faite sur la variable duale y au lieu de la variable primale x . Cela conduit à l'algorithme dual

$$\begin{cases} x_{n+1} = \text{prox}_{\tau G}(x_n - \tau K^* \bar{y}_n) \\ y_{n+1} = \text{prox}_{\sigma F^*}(y_n + \sigma K x_{n+1}) \\ \bar{y}_{n+1} = y_{n+1} + \theta (y_{n+1} - y_n). \end{cases}$$

On obtient alors dans ce cas un résultat similaire :

Théorème 16 *Supposons que le problème (5.13) admet une solution, notée (x^*, y^*) . Si on choisit $\tau > 0$, $\sigma > 0$ et $0 < \theta \leq 1$ vérifiant*

$$\max \left\{ \frac{1}{\tau\gamma + 1}, \frac{1}{\sigma\delta + 1} \right\} \leq \theta \leq \frac{1}{L^2\tau\sigma} \quad (5.40)$$

alors, pour tout ω tel que

$$\max \left\{ \frac{\theta + 1}{\tau\gamma + 2}, \frac{1}{\sigma\delta + 1} \right\} \leq \omega \leq \theta$$

alors on a la majoration suivante pour tout $N \in \mathbb{N}$ et pour tout $(x, y) \in X \times Y$:

$$\begin{aligned} 0 &\leq (1 - \omega L^2\tau\sigma) \frac{1}{2\tau} \|x_N - x\|^2 + \frac{1}{2\sigma} \|y_N - y\|^2 \\ &\quad + \sum_{n=1}^N \frac{1}{\omega^{n-1}} (\mathcal{L}(x_n, y) - \mathcal{L}(x, y_n)) \\ &\leq \frac{\omega^N}{2\tau} \|x_0 - x\|^2 + \frac{\omega^N}{2\sigma} \|y_0 - y\|^2. \end{aligned}$$

Posons pour tout $N \geq 1$

$$T_N = \sum_{n=1}^N \frac{1}{\omega^{n-1}} = \frac{1 - \omega^N}{\omega^{N-1}(1 - \omega)}$$

et définissons

$$X_N = \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} x_n \quad \text{et} \quad Y_N = \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega^{n-1}} y_n.$$

Alors on a l'encadrement suivant pour tout $(x, y) \in X \times Y$

$$\begin{aligned} 0 &\leq \frac{1 - \omega}{\omega(1 - \omega^N)} (1 - \omega L^2\tau\sigma) \frac{1}{2\tau} \|x - x_N\|^2 + \frac{1 - \omega}{\omega(1 - \omega^N)} \frac{1}{2\sigma} \|y - y_N\|^2 \\ &\quad + \mathcal{L}(X_N, y) - \mathcal{L}(x, Y_N) \\ &\leq \frac{1}{T_N} \frac{1}{2\tau} \|x - x_0\|^2 + \frac{1}{T_N} \frac{1}{2\sigma} \|y - y_0\|^2. \end{aligned}$$

5.3 Application à l'ADMM

On commence par rappeler que les itérations de l'ADMM sont équivalentes aux itérations de l'algorithme PDHG avec sur-relaxation étudié dans la section précédente, mais appliqué à un autre problème. Ainsi, on pourra utiliser les résultats de convergence pour en déduire le taux de convergence ergodique optimal d'ADMM. Par ailleurs, on remarquera qu'une légère relaxation d'un des pas dans l'ADMM permet théoriquement d'en accélérer la convergence.

5.3.1 Lien entre PDHG avec sur-relaxation et ADMM

On rappelle que les itérations de l'algorithme ADMM sont données par

$$\begin{cases} z_0 \in X \\ y_0 \in Y \end{cases} \quad \text{et} \quad \forall n \in \mathbb{N}, \quad \begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} \left\{ G(x) + \langle Kx, y_n \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 \right\} \\ z_{n+1} = \operatorname{argmin}_{z \in X} \left\{ F(z) - \langle z, y_n \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z\|^2 \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda} (Kx_{n+1} - z_{n+1}). \end{cases}$$

Posons $\xi_{n+1} = Kx_{n+1}$ et introduisons la fonction convexe

$$G_K(\xi) = \inf_{x \in X, Kx = \xi} G(x).$$

Par optimalité, on a pour tout $x \in X$

$$G(x_{n+1}) + \langle Kx_{n+1}, y_n \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z_n\|^2 \leq G(x) + \langle Kx, y_n \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2.$$

On en déduit en particulier que, si $Kx = Kx_{n+1}$, alors

$$G(x_{n+1}) + \langle Kx_{n+1}, y_n \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z_n\|^2 \leq G(x) + \langle Kx_{n+1}, y_n \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z_n\|^2.$$

Il s'ensuit que

$$\begin{aligned} G(x_{n+1}) + \langle Kx_{n+1}, y_n \rangle + \frac{1}{2\lambda} \|Kx_{n+1} - z_n\|^2 &= G_K(\xi_{n+1}) + \langle \xi_{n+1}, y_n \rangle + \frac{1}{2\lambda} \|\xi_{n+1} - z_n\|^2 \\ &= \min_{\xi \in Y} \left\{ G_K(\xi) + \langle \xi, y_n \rangle + \frac{1}{2\lambda} \|\xi - z_n\|^2 \right\}. \end{aligned}$$

Il vient en particulier que

$$\xi_{n+1} = \operatorname{argmin}_{\xi \in Y} \left\{ G_K(\xi) + \langle \xi, y_n \rangle + \frac{1}{2\lambda} \|\xi - z_n\|^2 \right\} = \operatorname{prox}_{\lambda G_K}(z_n - \lambda y_n).$$

La montée de gradient en y s'écrit alors

$$\begin{aligned} y_{n+1} = y_n + \frac{1}{\lambda} (\xi_{n+1} - z_{n+1}) &\iff z_{n+1} - \lambda y_{n+1} = \xi_{n+1} - \lambda (2y_{n+1} - y_n) \\ &\iff \xi_{n+1} + \lambda y_n = z_{n+1} + \lambda y_{n+1}, \end{aligned}$$

et implique d'une part que

$$\xi_{n+1} = \operatorname{prox}_{\lambda G_K}(\xi_n - \lambda (2y_n - y_{n-1}))$$

et d'autre part, en utilisant l'identité de MOREAU, que

$$z_{n+1} = \operatorname{prox}_{\lambda F}(\xi_{n+1} + \lambda y_n) = z_{n+1} + \lambda y_{n+1} - \lambda \operatorname{prox}_{F^*/\lambda}(y_n + \xi_{n+1}/\lambda).$$

On peut alors réécrire les mises-à-jour de x_{n+1} , y_{n+1} et z_{n+1} de la manière suivante

$$\begin{cases} \xi_0 \in Y^* \\ y_0 \in Y \\ \bar{y}_0 = y_0 \end{cases} \quad \text{et} \quad \forall n \in \mathbb{N}, \quad \begin{cases} \xi_{n+1} = \text{prox}_{\lambda G_K}(\xi_n - \lambda \bar{y}_n) \\ y_{n+1} = \text{prox}_{F^*/\lambda}(y_n + \xi_{n+1}/\lambda) \\ \bar{y}_{n+1} = y_{n+1} + (y_{n+1} - y_n). \end{cases} \quad (5.41)$$

D'après la section précédente, il s'agit de l'algorithme PDHG avec sur-relaxation sur la variable duale, qui résout le problème de recherche de point-selle

$$\min_{\xi \in X} \sup_{y \in Y} \left\{ G_K(\xi) + \langle \xi, y \rangle - F^*(y) \right\}$$

où on a $g = G_K$, $f = F$ et $A = \text{Id}$ (de norme $\tilde{L} = 1$). Le paramètre de relaxation vaut $\theta = 1$, les pas de temps valent respectivement $\tau = \lambda$ et $\sigma = 1/\lambda$. On posera par ailleurs $\tilde{\mathcal{L}}(\xi, y) = G_K(\xi) + \langle \xi, y \rangle - F^*(y)$.

5.3.2 Taux de convergence de l'ADMM classique

Utilisons le théorème 16 pour calculer le taux de convergence optimal de l'algorithme (5.41). Pour pouvoir l'appliquer, il nous faut tout d'abord vérifier que g est fortement convexe. Or, si G est fortement convexe (de paramètre γ), alors G_K est fortement convexe, de paramètre γ/L^2 . En effet, il est suffisant de montrer G_K^* est différentiable et que son gradient ∇G_K^* est lipschitzien de constante L^2/γ (voir chapitre 1). C'est bien le cas ici car

$$\begin{aligned} G_K^*(y + h) &= G^*(K^*(y + h)) = G^*(K^*y + K^*h) = G^*(K^*y) + \langle \nabla G^*(K^*y), K^*h \rangle \\ &\quad + o(\|K^*h\|). \end{aligned}$$

Par conséquent, g est fortement convexe, de paramètre $\tilde{\gamma} = \gamma/L^2$. On a par ailleurs par hypothèse que f^* est fortement convexe, de paramètre δ . On pose dans ce cas le conditionnement $\kappa = (1/\delta)/\tilde{\gamma} = L^2/(\gamma\delta)$.

On peut ainsi appliquer le théorème 16. La condition (5.40) est toujours vérifiée, puisque

$$\max \left\{ \frac{1}{\lambda\gamma/L^2 + 1}, \frac{1}{\delta/\lambda + 1} \right\} \leq 1.$$

Ainsi, pour tout $\lambda > 0$, Le taux de convergence vérifie

$$\max \left\{ \frac{1}{(\lambda\gamma)/(2L^2) + 1}, \frac{1}{\delta/\lambda + 1} \right\} \leq \omega \leq 1.$$

La borne inférieure de ce taux vaut $1/((\lambda\gamma)/(2L^2) + 1)$ lorsque $\lambda \leq \sqrt{2\delta L^2/\gamma}$ et vaut $1/(\delta/\lambda + 1)$ sinon. Le taux optimal est donc donné par

$$\omega^* = \frac{1}{\sqrt{(\gamma\delta)/(2L^2) + 1}} = \frac{1}{\sqrt{1/(2\kappa) + 1}} = \tilde{\omega} \quad \text{atteint pour} \quad \lambda = \sqrt{\frac{2\delta L^2}{\gamma}}.$$

Ainsi, si on pose $\Xi_N = \frac{1}{T_N} \sum_{n=1}^N \frac{1}{\omega_{n-1}} \xi_n$, on a pour tout $(\xi, y) \in Y^* \times Y$

$$\begin{aligned} 0 &\leq \frac{1 - \omega^*}{\omega^*(1 - (\omega^*)^N)} (1 - \omega^*) \sqrt{\frac{\gamma}{8\delta L^2}} \|\xi - \xi_N\|^2 + \frac{1 - \omega^*}{\omega^*(1 - (\omega^*)^N)} \sqrt{\frac{\delta L^2}{2\gamma}} \|y - y_N\|^2 \\ &\quad + \tilde{\mathcal{L}}(\Xi_N, y) - \tilde{\mathcal{L}}(\xi, Y_N) \\ &\leq \frac{1}{T_N} \sqrt{\frac{\gamma}{8\delta L^2}} \|\xi - \xi_0\|^2 + \frac{1}{T_N} \sqrt{\frac{\delta L^2}{2\gamma}} \|y - y_0\|^2 \end{aligned}$$

avec $T_N = (1 - (\omega^*)^N)/((\omega^*)^{N-1}(1 - \omega^*))$, qui s'écrit également pour tout $\xi = Kx$

$$\begin{aligned} 0 &\leq \frac{\gamma/L^2}{4(\sqrt{(\gamma\delta)/(2L^2)} + 1)} \|Kx - Kx_N\|^2 + \frac{\delta}{2\sqrt{(\gamma\delta)/(2L^2)}} \|y - y_N\|^2 \\ &\quad + \frac{\omega^*(1 - (\omega^*)^N)}{1 - \omega^*} (\tilde{\mathcal{L}}(KX_N, y) - \tilde{\mathcal{L}}(Kx, Y_N)) \\ &\leq (\omega^*)^N \left(\sqrt{\frac{\gamma}{8\delta L^2}} \|Kx - Kx_0\|^2 + \sqrt{\frac{\delta L^2}{2\gamma}} \|y - y_0\|^2 \right). \end{aligned}$$

Cet encadrement assure la convergence linéaire des itérées y_n et Kx_n lorsqu'elle est appliquée à $(x, y) = (x^*, y^*)$ (mais non nécessairement celle de x_n seul). À nouveau, il est possible, en choisissant de perdre le contrôle sur le *primal-dual gap*, d'obtenir comme taux de convergence $\tilde{\omega}$ à la place de ω^* . Par ailleurs, si on pose $E = G + F(K \cdot)$ et $\xi = Kx$, on a par définition,

$$E(x) = G_K(\xi) + F(\xi) = \sup_{y \in Y} \tilde{\mathcal{L}}(\xi, y) = \sup_{y \in Y} \tilde{\mathcal{L}}(Kx, y)$$

On en déduit que $E(x^*) = \tilde{\mathcal{L}}(Kx^*, y^*) \geq \tilde{\mathcal{L}}(Kx^*, Y_N)$, ce qui implique que l'inégalité précédente devient lorsque $x = x^*$ (en ignorant les deux premiers termes quadratiques)

$$\begin{aligned} 0 &\leq \frac{\omega^*(1 - (\omega^*)^N)}{1 - \omega^*} (\tilde{\mathcal{L}}(KX_N, y) - E(x^*)) \\ &\leq (\omega^*)^N \left(\sqrt{\frac{\gamma}{8\delta L^2}} \|Kx - Kx_0\|^2 + \sqrt{\frac{\delta L^2}{2\gamma}} \|y - y_0\|^2 \right). \end{aligned}$$

Si on connaît par ailleurs un ouvert borné \mathcal{B}_Y contenant y^* , alors on peut prendre la borne supérieure sur $y \in \mathcal{B}_Y$ dans cet encadrement, et obtenir

$$\begin{aligned} 0 &\leq \frac{\omega^*(1 - (\omega^*)^N)}{1 - \omega^*} (E(X_N) - E(x^*)) \\ &\leq (\omega^*)^N \left(\sqrt{\frac{\gamma}{8\delta L^2}} \|Kx^* - Kx_0\|^2 + \sqrt{\frac{\delta L^2}{2\gamma}} \sup_{y \in \mathcal{B}_Y} \|y - y_0\|^2 \right) \end{aligned}$$

ce qui implique une convergence linéaire (ergodique) de l'énergie primale, de taux ω^* .

5.3.3 Variante proposée de l'ADMM

On se propose maintenant de relaxer le choix du pas λ dans les mises-à-jour z et de y dans les itérations de l'ADMM. Plus précisément, si λ est remplacé par $\lambda' \leq \lambda$ dans chacune de ces deux mises-à-jour :

$$\begin{cases} x_{n+1} = \operatorname{argmin}_{x \in X} \left\{ G(x) + \langle Kx, y_n \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 \right\} \\ z_{n+1} = \operatorname{argmin}_{z \in X} \left\{ F(z) - \langle z, y_n \rangle + \frac{1}{2\lambda'} \|Kx_{n+1} - z\|^2 \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda'} (Kx_{n+1} - z_{n+1}) \end{cases}$$

alors on peut montrer que ces nouvelles itérations sont équivalentes à celles de l'algorithme suivant

$$\begin{cases} \xi^0 \in Y^* \\ y^0 \in Y \\ \bar{y}_0 = y_0 \end{cases} \quad \text{et} \quad \forall n \in \mathbb{N}, \quad \begin{cases} \xi^{n+1} = \operatorname{prox}_{\lambda G_K} (\xi^n - \lambda \bar{y}^n) \\ y^{n+1} = \operatorname{prox}_{F^*/\lambda'} (y^n + \xi^{n+1}/\lambda') \\ \bar{y}^{n+1} = y^{n+1} + \frac{\lambda'}{\lambda} (y^{n+1} - y^n) \end{cases}$$

où le paramètre de sur-relaxation vaut cette fois $\theta = \lambda'/\lambda \leq 1$. On a alors $\tau = \lambda$ et $\sigma = 1/\lambda'$. Déterminons le meilleur taux de convergence que cet algorithme peut atteindre, afin d'en déduire la valeur optimale des paramètres.

D'après le théorème 16, les pas λ et λ' doivent être liés par la relation

$$\max \left\{ \frac{1}{\lambda\gamma/L^2 + 1}, \frac{1}{\delta/\lambda' + 1} \right\} \leq \frac{\lambda'}{\lambda} \leq 1.$$

Fixons $\lambda' > 0$. Cherchons les conditions que doit vérifier λ pour que l'encadrement précédent soit valable. De manière équivalente, on cherche λ satisfaisant au moins un des deux ensembles de conditions suivants

$$\frac{1}{\lambda\gamma/L^2 + 1} \leq \frac{1}{\delta/\lambda' + 1} \quad \text{et} \quad \frac{1}{\delta/\lambda' + 1} \leq \frac{\lambda'}{\lambda} \quad \text{et} \quad \frac{\lambda'}{\lambda} \leq 1 \quad (5.42)$$

ou

$$\frac{1}{\lambda\gamma/L^2 + 1} \geq \frac{1}{\delta/\lambda' + 1} \quad \text{et} \quad \frac{1}{\lambda\gamma/L^2 + 1} \leq \frac{\lambda'}{\lambda} \quad \text{et} \quad \frac{\lambda'}{\lambda} \leq 1. \quad (5.43)$$

Suivant les valeurs de λ' , on va établir si ces deux ensembles de conditions admettent des éléments. Commençons par les conditions (5.42), qui sont équivalentes à

$$\max \left\{ \frac{\delta L^2}{\gamma \lambda'}, \lambda' \right\} \leq \lambda \leq \lambda' + \delta. \quad (5.44)$$

Pour que de tels λ existent, il faut (puisqu'on a nécessairement $\lambda' \leq \lambda' + \delta$) que

$$\frac{\delta L^2}{\gamma \lambda'} \leq \lambda' + \delta. \quad (5.45)$$

Il est aisé de vérifier que tout réel positif λ' vérifiant (5.45) est supérieur à $(\lambda')^*$ (strictement si on considère également l'inégalité stricte dans (5.45)), donné par

$$(\lambda')^* = \frac{\delta}{2} \left(\sqrt{1 + \frac{4L^2}{\gamma\delta}} - 1 \right).$$

Aussi, l'ensemble des réels λ strictement positifs vérifiant les conditions (5.42) est un intervalle non vide si et seulement si $\lambda' \geq (\lambda')^*$. La borne inférieure est alors donnée par λ' si $\lambda' > \sqrt{\delta L^2/\gamma}$ et par $(\delta L^2)/(\gamma\lambda')$ sinon. Intéressons-nous à présent aux conditions (5.43). On peut vérifier qu'elles s'écrivent de manière équivalente

$$\lambda' \leq \lambda \leq \frac{\delta L^2}{\gamma\lambda'} \quad \text{et} \quad \lambda \left(1 - \frac{\gamma\lambda'}{L^2} \right) \leq \lambda'. \quad (5.46)$$

Le premier encadrement entraîne que $\lambda' \leq (\delta L^2)/(\gamma\lambda')$, soit $\lambda' \leq \sqrt{\delta L^2/\gamma}$. Aussi, on peut d'ores et déjà affirmer que, si λ' ne vérifie pas cette inégalité, alors aucun réel λ ne peut satisfaire les conditions (5.43). On supposera donc dans la suite de cette analyse que le premier encadrement peut être satisfait. Pour interpréter la seconde relation, il est nécessaire de tenir compte du signe du terme entre parenthèses. Deux cas de figure sont envisageables. Si ce terme est négatif ou nul, c'est-à-dire si $\lambda' \geq L^2/\gamma$, alors cette inégalité est toujours vraie. Les conditions (5.46) impliquent alors

$$\frac{L^2}{\gamma} \leq \lambda' \leq \sqrt{\frac{\delta L^2}{\gamma}}.$$

Ainsi, si $L^2/\gamma \leq \delta$, alors il existe des réels λ vérifiant les conditions (5.43) lorsque λ' vérifie l'encadrement précédent. Si $L^2/\gamma > \delta$ ou si $\lambda' < L^2/\gamma$, alors $1 - \gamma\lambda'/L^2$ est positif strictement, et les conditions (5.46) deviennent

$$\lambda' \leq \lambda \leq \min \left\{ \frac{\delta L^2}{\gamma\lambda'}, \frac{\lambda'}{1 - (\gamma\lambda')/L^2} \right\}.$$

Notons que, puisque L^2/γ vérifie (strictement) la relation (5.45), alors il est strictement supérieur à $(\lambda')^*$. On montre de même que $\sqrt{\delta L^2/\gamma} > (\lambda')^*$. Calculons ensuite le membre de droite de l'encadrement précédent. Puisque

$$\frac{\delta L^2}{\gamma\lambda'} \leq \frac{\lambda'}{1 - (\gamma\lambda')/L^2} \quad \iff \quad 0 \leq \frac{\gamma}{L^2}(\lambda')^2 + \frac{\gamma\delta}{L^2}\lambda' - \delta \quad \iff \quad \lambda' \geq (\lambda')^*$$

on en déduit que, si $\lambda' < (\lambda')^*$, alors le pas λ est encadré par

$$\lambda' \leq \lambda \leq \frac{\lambda'}{1 - (\gamma\lambda')/L^2}. \quad (5.47)$$

Si $\sqrt{\delta L^2/\gamma} \geq \lambda' \geq (\lambda')^*$, nous ignorons la seconde condition dans (5.46) et λ est contraint par le premier encadrement. Reprenons à présent l'analyse complète.

1. On a montré que, si $\lambda' < (\lambda')^*$, alors les conditions (5.42) ne peuvent être satisfaites. En revanche, les conditions (5.43) conduisent dans ce cas à l'encadrement (5.47). On a en effet $\lambda' < \sqrt{\delta L^2/\gamma}$ et $\lambda' < L^2/\gamma$.

2. On a ensuite établi que, si $\lambda' > \sqrt{\delta L^2/\gamma}$, alors ce sont cette fois-ci les conditions (5.43) qui ne peuvent être satisfaites. Néanmoins, cette hypothèse implique que $\lambda' \geq (\lambda')^*$; les conditions (5.42) prennent alors la forme (5.44), où la borne inférieure vaut λ' :

$$\lambda' \leq \lambda \leq \lambda' + \delta. \quad (5.48)$$

3. Enfin, si λ' est compris entre $(\lambda')^*$ et $\sqrt{\delta L^2/\gamma}$, alors les deux ensembles de conditions (5.42) et (5.43) peuvent être satisfaites. Elles s'écrivent respectivement

$$\frac{\delta L^2}{\gamma \lambda'} \leq \lambda \leq \lambda' + \delta \quad \text{et} \quad \lambda' \leq \lambda \leq \frac{\delta L^2}{\gamma \lambda'}$$

ces deux encadrements pouvant être réunis en un seul, donné par (5.48).

On voit donc finalement que les conditions sur λ ne dépendent que de la valeur relative de λ' par rapport à $(\lambda')^*$.

Maintenant qu'on a établi les valeurs admissibles pour les pas λ et λ' , on peut minimiser le taux de convergence ω . On rappelle qu'il est contraint par

$$\max \left\{ \frac{\lambda'/\lambda + 1}{\lambda\gamma/L^2 + 2}, \frac{1}{\delta/\lambda' + 1} \right\} \leq \omega \leq \frac{\lambda'}{\lambda}.$$

On cherche donc à minimiser la borne inférieure de cet encadrement. Calculons le minimum de la fonction $\lambda \mapsto (\lambda'/\lambda + 1)(\lambda\gamma/L^2 + 2)$ pour λ satisfaisant la contrainte (5.47) si $\lambda' < (\lambda')^*$ et la contrainte (5.48) sinon. Puisqu'il s'agit d'une fonction strictement décroissante, son minimum est atteint lorsque λ est maximal. Il s'ensuit que

$$\min_{\substack{\lambda \text{ satisfaisant} \\ (5.48) \text{ et } (5.47)}} \left\{ \frac{\lambda'/\lambda + 1}{\lambda\gamma/L^2 + 2} \right\} = \begin{cases} 1 - \frac{\gamma\lambda'}{L^2} & \text{si } \lambda' < (\lambda')^* \\ \frac{1}{\delta/\lambda' + 1} \frac{\delta/\lambda' + 2}{(\delta/\lambda' + 1)\gamma\lambda'/L^2 + 2} & \text{si } \lambda' \geq (\lambda')^*. \end{cases}$$

Comparons maintenant ce minimum à $1/(\delta/\lambda' + 1)$. Pour ce faire, on montre d'une part que

$$1 - \frac{\gamma\lambda'}{L^2} \geq \frac{1}{\delta/\lambda' + 1} \iff \frac{\delta}{\lambda'} - \frac{\delta}{\lambda'} \frac{\gamma\lambda'}{L^2} - \frac{\gamma\lambda'}{L^2} \geq 0 \iff \lambda' \leq (\lambda')^*$$

et d'autre part que

$$\begin{aligned} \frac{1}{\delta/\lambda' + 1} \frac{\delta/\lambda' + 2}{(\delta/\lambda' + 1)\gamma\lambda'/L^2 + 2} \leq \frac{1}{\delta/\lambda' + 1} & \iff \frac{\delta}{\lambda'} \leq \left(\frac{\delta}{\lambda'} + 1 \right) \frac{\gamma\lambda'}{L^2} \\ & \iff \lambda' \geq (\lambda')^*. \end{aligned}$$

Nous avons donc prouvé que ω est minoré par

$$\omega^*(\lambda') = \begin{cases} 1 - \frac{\gamma\lambda'}{L^2} & \text{si } \lambda' < (\lambda')^* \\ \frac{1}{\delta/\lambda' + 1} & \text{si } \lambda' \geq (\lambda')^* \end{cases}$$

qui est alors minimal pour $\lambda' = (\lambda')^*$. On en déduit le taux

$$\omega^* = \frac{\sqrt{1 + (4L^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4L^2)/(\gamma\delta)} + 1} = \frac{\sqrt{1 + 4\kappa} - 1}{\sqrt{1 + 4\kappa} + 1}$$

obtenu lorsque les pas de temps λ et λ' sont choisis comme suit :

$$\lambda' = \frac{\delta}{2} \left(\sqrt{1 + \frac{4L^2}{\gamma\delta}} - 1 \right) \quad \text{et} \quad \lambda = \lambda' + \delta = \frac{\delta}{2} \left(\sqrt{1 + \frac{4L^2}{\gamma\delta}} + 1 \right).$$

Le lecteur pourra vérifier que cette valeur est strictement inférieure au taux ω^* obtenu dans le cas de l'ADMM classique. Finalement, en reprenant les mêmes calculs que pour l'ADMM classique, on montre que pour tout $N \geq 1$

$$\begin{aligned} 0 &\leq \frac{\lambda'}{2} \|y - y_N\|^2 + \frac{\omega^*(1 - (\omega^*)^N)}{1 - \omega^*} \left(\tilde{\mathcal{L}}(\Xi_N, y) - \tilde{\mathcal{L}}(\xi, Y_N) \right) \\ &\leq (\omega^*)^N \left(\frac{1}{2\lambda} \|\xi - \xi_0\|^2 + \frac{\lambda'}{2} \|y - y_0\|^2 \right) \end{aligned}$$

car, ici, les paramètres optimaux vérifient $\omega^* = \theta = 1/(\tau\sigma)$. On voit ainsi que choisir les paramètres λ et λ' de manière à obtenir la plus petite valeur de ω conduit à perdre le contrôle sur la convergence de la variable x . Cette convergence reste néanmoins assurée par la forte convexité (cf. proposition 7). Pour la convergence de la variable duale, le taux

$$\tilde{\omega} = \frac{\sqrt{1 + (4L^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4L^2)/(\gamma\delta)} + 3} = \frac{\sqrt{1 + 4\kappa} - 1}{\sqrt{1 + 4\kappa} + 3}$$

est à nouveau valable.

REMARQUE : Si l'on tient à conserver ce contrôle, une manière simple de procéder est de considérer un $\tilde{L} \neq L$ et de poser

$$\lambda' = \frac{\delta}{2} \left(\sqrt{1 + \frac{4\tilde{L}^2}{\gamma\delta}} - 1 \right) \quad \text{et} \quad \lambda \geq \left(\frac{L}{\tilde{L}} \right)^2 \frac{\delta}{2} \left(\sqrt{1 + \frac{4\tilde{L}^2}{\gamma\delta}} + 1 \right)$$

de sorte d'avoir $\lambda\gamma/L^2 + 1 \geq \delta/\lambda' + 1$. La condition (5.40) du théorème 16 s'écrit alors

$$\frac{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} + 1} \leq \left(\frac{\tilde{L}}{L} \right)^2 \frac{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} + 1} \leq 1$$

On en déduit \tilde{L} doit être compris entre L et $L\sqrt{1 + \sqrt{\gamma\delta/L^2}}$. Sous cette condition, le théorème 16 assure que le taux de convergence ω est minoré par

$$\max \left\{ \frac{\lambda'/\lambda + 1}{\lambda\gamma/L^2 + 2}, \frac{1}{\delta/\lambda + 1} \right\}$$

dont on montre qu'il vaut dans le cas présent $1/(\delta/\lambda + 1)$. Ainsi, le meilleur taux est donné par

$$\omega^* = \frac{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} - 1}{\sqrt{1 + (4\tilde{L}^2)/(\gamma\delta)} + 1} \quad \text{avec} \quad 1 - \frac{\omega^*}{\theta} = 1 - \left(\frac{L}{\tilde{L}} \right)^2 \leq \frac{\sqrt{\gamma\delta/L^2}}{1 + \sqrt{\gamma\delta/L^2}}$$

et on a pour tout $N \geq 1$

$$\begin{aligned} 0 &\leq \frac{1 - (L/\tilde{L})^2}{2\lambda} \|\xi - \xi_N\|^2 + \frac{\lambda'}{2} \|y - y_N\|^2 + \frac{\omega^*(1 - (\omega^*)^N)}{1 - \omega^*} \left(\tilde{\mathcal{L}}(\Xi_N, y) - \tilde{\mathcal{L}}(\xi, Y_N) \right) \\ &\leq (\omega^*)^N \left(\frac{1}{2\lambda} \|\xi - \xi_0\|^2 + \frac{\lambda'}{2} \|y - y_0\|^2 \right). \end{aligned}$$

5.4 Exemples numériques

5.4.1 Algorithme FISTA

On peut comparer les taux obtenus pour l'ADMM modifié avec ceux obtenus pour un autre algorithme de descente proximale accéléré, proposé par NESTEROV [11] et BECK et TEBoulLE [2]. Il s'agit de considérer la forme sans contrainte du problème primal, donnée par

$$\min_{x \in X} \left\{ F(Kx) + G(x) \right\}$$

et de le résoudre par méthode de gradient explicite-implicite (voir chapitre 1), en ajoutant un pas de sur-relaxation :

$$x_0 = \bar{x}_0 \in X \quad \text{et} \quad \begin{cases} x_{k+1} = \text{prox}_{\tau G} \left(\bar{x}_k - \tau K^* \nabla \left(F(K\bar{x}_k) \right) \right) \\ \bar{x}_{k+1} = x_{k+1} + \beta_{k+1} (x_{k+1} - x_k) \end{cases}$$

où les paramètres variables de la sur-relaxation sont donnés par

$$t_{k+1} = \frac{1 - qt_k^2 + \sqrt{(1 - qt_k^2)^2 + 4t_k^2}}{2} \quad \text{et} \quad \beta_k = (1 + \tau\gamma - t_{k+1}\tau\gamma) \frac{t_k - 1}{t_{k+1}}$$

avec $q = \tau\gamma/(1 + \tau\gamma)$, pour $\tau \in]0; \delta/L^2]$. Dans le cas où F^* et G sont supposées fortement convexes, de modules respectifs δ et γ , on peut montrer [6] que la convergence de cet algorithme est linéaire. Lorsque les pas sont choisis constants ($t_k = t = 1/\sqrt{\tau\gamma/(1 + \tau\gamma)}$), ce taux vaut alors $1 - \sqrt{q}$ [6, Remark B.2], ce qui conduit au taux minimal

$$\omega_{\text{FISTA}}^* = 1 - \sqrt{\frac{\tau\gamma}{1 + \tau\gamma}} = 1 - \sqrt{\frac{\gamma\delta/L^2}{1 + \gamma\delta/L^2}} = 1 - \sqrt{\frac{1}{1 + \kappa}}$$

et on a la majoration suivante pour tout $k \in \mathbb{N}$

$$E(x_k) - E(x^k) \leq (\omega_{\text{FISTA}}^*)^k \left(E(x_0) - E(x^k) + \gamma \frac{\|x_0 - x^*\|^2}{2} \right).$$

L'intérêt de cet algorithme réside dans le cas où ∇F et prox_G sont calculables.

5.4.2 Comparaison théorique avec FISTA et PDHG

On peut comparer les taux théoriques obtenus pour chacun des quatre algorithmes considérés dans cette section, qui sont l'ADMM, l'ADMM modifié, PDHG sur-relaxé et FISTA (cas fortement convexe). Ainsi, la figure 5.1 affiche la valeur des taux théoriques minimaux en fonction du conditionnement. Ainsi, il apparaît que, comme prévu, l'ADMM classique affiche le taux le moins bon, que l'ADMM modifié et PDHG offrent le même taux, qui est légèrement moins bon que celui de FISTA. Tous les taux tendent vers 1 lorsque le problème est mal conditionné.

5.4.3 Premier test quadratique

On commence dans ce paragraphe par comparer numériquement les performances de l'ADMM et de l'ADMM modifié sur un exemple simple.

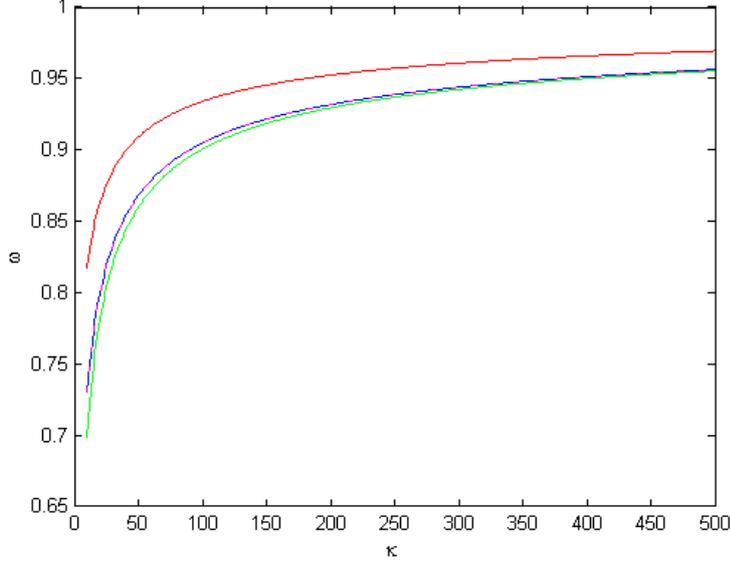


FIGURE 5.1 – Comparaison des taux de convergence ω théoriques en fonction du conditionnement κ , pour l'ADMM classique (rouge), l'ADMM modifié (magenta), PDHG sur-relaxé (bleu) et FISTA (vert). L'ordonnée sont en échelle logarithmique.

Problème considéré Soit N un entier naturel. Considérons le problème suivant sous contrainte égalité :

$$\min_{\substack{x=(x_i)_{i \in \llbracket 0; N-1 \rrbracket} \in \mathbb{R}^N \\ x_0=1}} \left\{ \frac{M-m}{2} \|K_N x\|^2 + \frac{m}{2} \|x\|^2 \right\} \quad (5.49)$$

où l'opérateur linéaire $K_N : \mathbb{R}^N \rightarrow \mathbb{R}^{N-1}$ est défini par $(K_N x)_i = (x_{i+1} - x_i)/2$ pour tout $i \in \llbracket 0; N-2 \rrbracket$, avec $\|K_N\| \leq 1$. Le conditionnement de ce problème vaut M/m , de sorte que, si l'on choisit M très grand devant m , alors ce problème est mal conditionné. On peut donc poser $F(z) = (M-m)\|z\|^2/2$ pour tout $z \in \mathbb{R}^{N-1}$ et $G(x) = m\|x\|^2/2 + \chi_{\{1\}}(x_0)$ pour tout $x = (x_i)_{i \in \llbracket 0; N-1 \rrbracket} \in \mathbb{R}^N$. Ces deux fonctions sont fortement convexes, de modules respectifs $M-m$ et $\gamma = m$. On vérifie que $F^*(y) = \|y\|^2/(M-m)/2$, qui est fortement convexe, de module $\delta = 1/(M-m)$. Appliquons à ce problème les deux versions de l'ADMM présentées dans cette section.

Résolution explicite Le minimiseur du problème (5.49) peut être calculé de manière explicite. Il suffit pour cela d'introduire le sous-vecteur \tilde{x} de x , défini comme suit :

$$\forall i \in \llbracket 0; N-2 \rrbracket, \quad \tilde{x}_i = x_{i+1}$$

de sorte que $x = (1, \tilde{x})$. Le problème sous contrainte (5.49) peut alors s'écrire sous la forme non contrainte

$$\min_{\tilde{x}=(\tilde{x}_i)_{i \in \llbracket 0; N-2 \rrbracket} \in \mathbb{R}^{N-1}} \left\{ \frac{M-m}{2} \left(\|K_{N-1} \tilde{x}\|^2 + \frac{(\tilde{x}_0 - 1)^2}{4} \right) + \frac{m}{2} (\|\tilde{x}\|^2 + 1) \right\}$$

en remarquant que $(K_N x) = K_{N-1} \tilde{x}$. Le minimiseur \tilde{x}^* de ce problème fortement convexe est donné par l'équation d'EULER, qui assure que $\tilde{x}^* = A^{-1}b$, avec

$$A = m I_{N-1} + (M-m) {}^t K_{N-1} K_{N-1} + \frac{M-m}{4} e_{0,0}$$

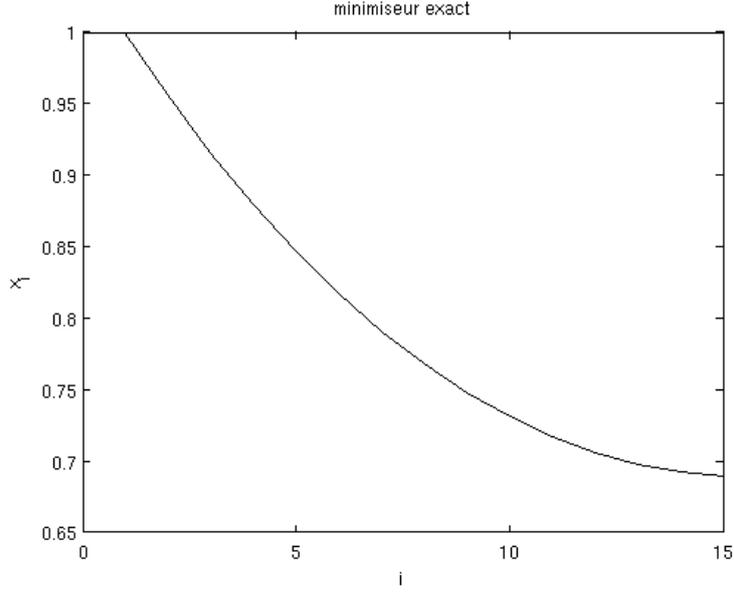


FIGURE 5.2 – Minimiseur théorique du problème (5.49).

avec $e_{0,0}$ la matrice carrée de taille $N-1$ ayant tous ses coefficients nuls à l'exception du coefficient d'indice $(0,0)$, qui vaut 1, et $b = (M-m)e_0/4$, où e_0 est le premier vecteur de la base canonique de \mathbb{R}^{N-1} . Le minimiseur du problème initial (5.49) est alors donné par $x^* = (1, \tilde{x}^*)$. Pour $M = 1000$ et $m = 1$, la figure 5.2 affiche l'allure de x^* .

Application de l'ADMM L'ADMM modifié s'écrit pour ce problème

$$\begin{cases} x_{n+1} = \underset{\substack{x=(x_i)_{i \in \mathbb{R}^N} \\ x_0=1}}{\operatorname{argmin}} \left\{ \frac{m}{2} \|x\|^2 + \langle Kx, y_n \rangle + \frac{1}{2\lambda} \|Kx - z_n\|^2 \right\} \\ z_{n+1} = \underset{z \in \mathbb{R}^{N-1}}{\operatorname{argmin}} \left\{ \frac{M-m}{2} \|z\|^2 - \langle z, y_n \rangle + \frac{1}{2\lambda'} \|Kx_{n+1} - z\|^2 \right\} \\ y_{n+1} = y_n + \frac{1}{\lambda'} (Kx_{n+1} - z_{n+1}). \end{cases}$$

La mise-à-jour de z s'écrit grâce à l'équation d'EULER

$$z_{n+1} = \frac{y_n + Kx_{n+1}/\lambda'}{M-m+1/\lambda'}.$$

Pour calculer explicitement la mise-à-jour de x , on utilise à nouveau les sous-vecteurs introduits au paragraphe précédent. Le problème à résoudre s'écrit alors sans contrainte (en ignorant les termes constants)

$$\min_{\tilde{x}=(\tilde{x}_i)_{i \in \mathbb{R}^{N-1}}} \left\{ \frac{m}{2} \|\tilde{x}\|^2 + \langle \tilde{x}, {}^t K_{N-1} \tilde{y}_n \rangle + \frac{1}{2\lambda} \left(\|K_{N-1} \tilde{x} - \tilde{z}_n\|^2 + \left(\frac{\tilde{x}_0 - 1}{2} - (z_n)_0 \right)^2 \right) \right\}.$$

L'équation d'EULER assure cette fois que $\tilde{x}_{n+1} = A_n^{-1} b_n$, avec

$$A_n = m I_{N-1} + \frac{1}{\lambda} {}^t K_{N-1} K_{N-1} + \frac{1}{4\lambda} e_{0,0}$$

et
$$b_n = -{}^t K_{N-1} \tilde{y}_n + \frac{1}{\lambda} {}^t K_{N-1} \tilde{z}_n + \left(-\frac{(y_n)_0}{2} + \frac{1}{2\lambda} (z_n)_0 + \frac{1}{4\lambda} \right) e_0.$$

On a alors $x_{n+1} = (1, \tilde{x}_{n+1})$.

Choix des paramètres On teste dans cette expérience deux jeux de paramètres différents pour λ et λ' . Le premier (désigné sous le nom d'ADMM classique) correspond au paramètre optimal pour l'ADMM classique déterminé dans cette section, qui vaut dans ce cas

$$\lambda = \lambda' = \sqrt{\frac{2L^2}{m(M-m)}}.$$

On prendra $L = 1$. Le second jeu de paramètres (ADMM modifié) correspond aux paramètres optimaux dans l'ADMM modifié, conduisant au meilleur taux de convergence :

$$\lambda = \frac{1}{2(M-m)} \left(\sqrt{1 + \frac{4L^2(M-m)}{m}} + 1 \right)$$

et

$$\lambda' = \lambda - \frac{1}{M-m} = \frac{1}{2(M-m)} \left(\sqrt{1 + \frac{4L^2(M-m)}{m}} - 1 \right).$$

Mesure de la convergence Pour comparer la convergence des trois jeux de paramètres, on utilisera trois outils :

1. la distance quadratique au minimiseur $\|x_n - x^*\|^2$;
2. la différence absolue entre l'énergie à l'itération n et l'énergie minimale ;
3. un *primal-dual gap*.

Concernant le dernier outil, le *gap* qui devrait être pris en compte est celui correspondant au problème primal-dual résolu, qui est la recherche de point-selle du lagrangien augmenté associé au problème initial (5.49). Malheureusement, l'énergie primale donnée par

$$\sup_{y \in \mathbb{R}^{N-1}} \left\{ \frac{m}{2} \|x_n\|^2 + \langle Kx_n - z_n, y_n \rangle + \frac{M-m}{2} \|z_n\|^2 + \frac{1}{2\lambda'} \|Kx_n - z_n\|^2 \right\}$$

est généralement infinie, car Kx_n n'est pas contraint à être égal à z_n . De ce fait, utiliser ce *gap* n'est pas pertinent. C'est pourquoi on se propose ici d'utiliser la conjuguée convexe de $x \mapsto F(Kx)$, ce qui permet de réécrire le problème primal (5.49) sous la forme primale-duale

$$\min_{\substack{x=(x_i)_{i \in \llbracket 0; N-1 \rrbracket} \in \mathbb{R}^N \\ x_0=1}} \sup_{z' \in \mathbb{R}^{N-1}} \left\{ \frac{m}{2} \|x\|^2 + \langle Kx, z' \rangle - \frac{1}{2(M-m)} \|z'\|^2 \right\}$$

Ce problème admet au moins une solution, donnée par $(x^*, (z')^*)$, avec

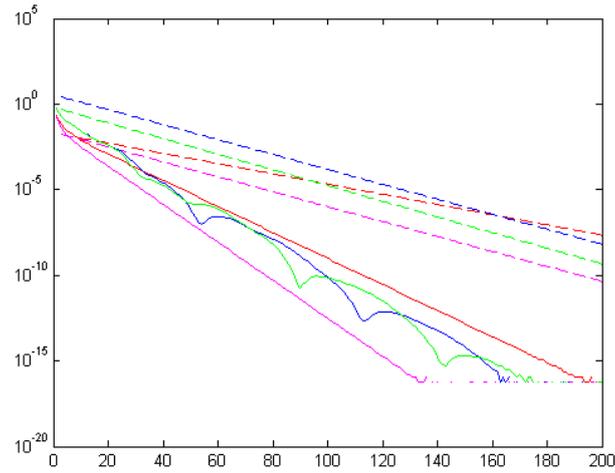
$$(z')^* = \operatorname{argmax}_{z' \in \mathbb{R}^{N-1}} \left\{ \frac{m}{2} \|x^*\|^2 + \langle Kx^*, z' \rangle - \frac{1}{2(M-m)} \|z'\|^2 \right\} = (M-m) Kx^*$$

(on résout l'équation d'EULER). Le *gap* pour ce problème vaut

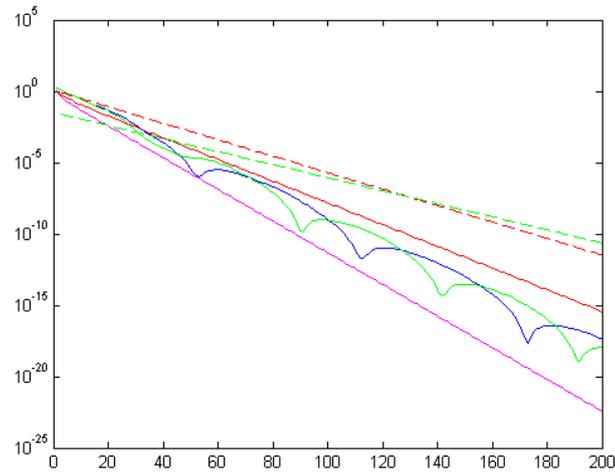
$$\mathcal{G}(x_n, z'_n) = \mathcal{P}(x_n) - \mathcal{D}(z'_n)$$

où l'énergie primale en x_n vaut $\mathcal{P}(x_n) = m \|x_n\|^2/2 + (M-m) \|Kx_n\|^2/2$, tandis que l'énergie duale en z'_n vaut

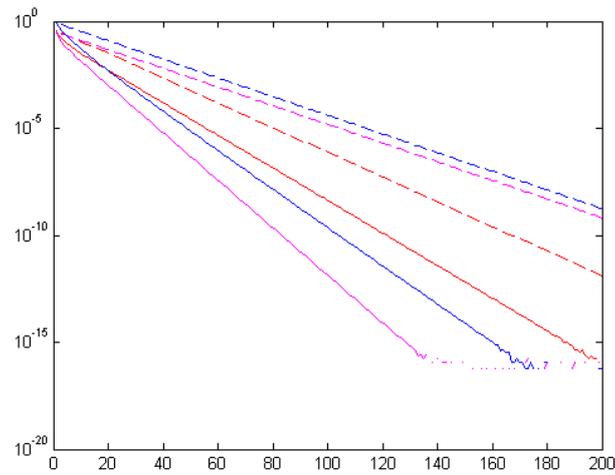
$$\mathcal{D}'(z'_n) = \frac{m}{2} \|x'\|^2 + \langle Kx', z'_n \rangle - \frac{1}{2(M-m)} \|z'_n\|^2$$



(a) Énergie $E(x_n) - E(x^*)$



(b) Distance au minimiseur $\|x_n - x^*\|^2$



(c) Primal-dual gap

FIGURE 5.3 – Mesure de la convergence pour l’ADMM (rouge), l’ADMM modifié (magenta), PDHG sur-relaxé (bleu) et FISTA (vert) pour le problème quadratique. En trait plein, les mesures empiriques, en pointillé, les valeurs théoriques (décroissance linéaire en ω pour l’énergie et le *gap* et en $\tilde{\omega}$ pour la convergence de la variable primale, sauf pour FISTA). Les ordonnées sont en échelle logarithmique.

avec $\tilde{x}' = -{}^t K_{N-1} z'_n / m$. On sait que le *gap* $\mathcal{G}(x_n; z'_n)$ doit tendre vers 0 si $(x_n)_n$ tend vers x^* et $(z'_n)_n$ tend vers $(M - m) K x^* = (M - m) z^*$. On en déduit en particulier que la suite $(\mathcal{G}(x_n; (M - m) z_n))_n$ doit tendre vers 0. C'est ce *gap*-là que l'on considèrera dans les expériences réalisées.

La figure 5.3 présente l'évolution des trois mesures décrites plus haut, pour un même nombre d'itérations. On affiche par ailleurs l'évolution de ces mêmes mesures obtenues lorsque l'on applique le PDHG sur-relaxé avec les paramètres optimaux déterminés au paragraphe 5.2.3 et lorsque l'on applique FISTA (cas fortement convexe) avec des pas constants (voir paragraphe précédent). Dans ce dernier cas, on ne considère que les deux premières quantités (le *gap* étant mal défini).

5.4.4 Application au débruitage TV-Huber

Problème considéré Soit $u^h \in \mathbb{R}^{3N_x N_y}$ une image couleur RGB (éventuellement bruitée). On cherche à résoudre

$$\min_{v^h \in \mathbb{R}^{3N_x N_y}} \left\{ \frac{1}{2\mu} \|v^h - u^h\|^2 + \frac{1}{2} \text{TV}_{\text{Huber}}^h(v^h) \right\} \quad (5.50)$$

où l'opérateur linéaire gradient $\nabla^h : \mathbb{R}^{3N_x N_y} \rightarrow \mathbb{R}^{3N_x N_y} \times \mathbb{R}^{3N_x N_y}$ est défini pour toute image v^h par $\nabla^h v^h = ({}^t(\delta_x^h v^h, \delta_y^h v^h))$, où les différences finies sont données pour tout indice $(i, j) \in \llbracket 0; N_x - 1 \rrbracket \times \llbracket 0; N_y - 1 \rrbracket$ par

$$(\delta_x^h v^h)_{i,j} = \begin{cases} v_{i+1,j}^h - v_{i,j}^h & \text{si } i < N_x - 1 \\ 0 & \text{sinon} \end{cases} \quad \text{et} \quad (\delta_y^h v^h)_{i,j} = \begin{cases} v_{i,j+1}^h - v_{i,j}^h & \text{si } j < N_y - 1 \\ 0 & \text{sinon} \end{cases}$$

et avec $\text{TV}_{\text{Huber}}^h(v^h)$ défini par

$$\text{TV}_{\text{Huber}}^h(v^h) = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} h(\|(\nabla^h v^h)_{i,j}\|) \quad \text{où} \quad h(t) = \begin{cases} \frac{|t|^2}{2} & \text{si } |t| \leq 1 \\ |t| - \frac{1}{2} & \text{si } |t| > 1. \end{cases}$$

Ce problème peut être interprété comme un problème de débruitage, avec un terme d'attache aux données quadratique et un terme de régularisation défini par une version régularisée de TV (voir figure 5.4), qui agit comme une régularisation quadratique lorsque les variations sont faibles et comme la régularisation TV sinon. Ce modèle peut donc être rapproché du modèle ROF étudié plus en détail dans le chapitre suivant. Le paramètre $\mu > 0$ permet de pondérer l'importance relative du terme d'attache aux données par rapport au terme de régularisation.

Application de l'ADMM Posons $G = \|\cdot - u^h\|^2 / (2\mu)$ et

$$F(\phi) = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} h(\|\phi_{i,j}\|) \quad \text{avec} \quad F^*(\xi) = \sum_{i=0}^{N_x-1} \sum_{j=0}^{N_y-1} \frac{1}{2} \|\xi_{i,j}\|^2 + \chi_{[0;1]}(\|\xi_{i,j}\|).$$

L'ADMM modifié s'écrit pour ce problème

$$\begin{cases} v_{n+1}^h = \underset{v^h \in \mathbb{R}^{3N_x N_y}}{\text{argmin}} \left\{ \frac{1}{2\mu} \|v^h - u^h\|^2 + \langle \nabla^h v^h, \xi_n^h \rangle + \frac{1}{2\lambda} \|\nabla^h v^h - \phi_n^h\|^2 \right\} \\ \phi_{n+1}^h = \underset{\phi^h \in (\mathbb{R}^{3N_x N_y})^2}{\text{argmin}} \left\{ F(\phi^h) - \langle \phi^h, \xi_n^h \rangle + \frac{1}{2\lambda'} \|\nabla^h v_{n+1}^h - \phi^h\|^2 \right\} \\ \xi_{n+1}^h = \xi_n^h + \frac{1}{\lambda'} (\nabla^h v_{n+1}^h - \phi_{n+1}^h). \end{cases}$$



(a) Image originale u^h (b) Image bruitée g^h (c) Image débruitée v^*

FIGURE 5.4 – Image originale u^h (à gauche), image bruitée g^h (au milieu) et image débruitée v^* (à droite). Le bruit ajouté est un bruit blanc gaussien additif, de variance 100 (les images sont à valeurs entre 0 et 255). L'image débruitée v^* est obtenue en appliquant une régularisation quadratique avec une attache aux données quadratique. Source : détail de l'image *Hepatica nobilis flowers*, par Archenzo.

Chacune des deux minimisations partielles se résolvent grâce à l'équation d'EULER, qui assure que la mise-à-jour de l'image v^h est calculée par

$$v_{n+1}^h = \left(\frac{1}{\mu} \mathbf{I} + \frac{1}{\lambda} (\nabla^h)^* \nabla^h \right)^{-1} \left(\frac{1}{\mu} u^h + \frac{1}{\lambda} (\nabla^h)^* \phi_n^h - (\nabla^h)^* \xi_n^h \right)$$

tandis que la mise-à-jour de la seconde variable primale ϕ^h est donnée par

$$\phi_{n+1}^h = \frac{\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}}{|\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}|} |(\phi_{n+1})_{i,j}|$$

avec

$$|(\phi_{n+1})_{i,j}| = \begin{cases} \frac{\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}}{\tau' + 1} & \text{si } |\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}| \leq \tau' + 1 \\ |\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}| - \tau' & \text{si } |\tau'(\xi_n)_{i,j} + (\nabla v_{n+1})_{i,j}| > \tau' + 1. \end{cases} \quad (5.51)$$

Choix des paramètres On va à nouveau tester les deux jeux de paramètres (λ, λ') correspondant à l'ADMM classique et à l'ADMM modifié.

Commençons par déterminer la régularité du problème. Les fonctions F^* et G sont fortement convexes, de modules respectifs $\delta = 1$ et $\gamma = 1/\mu$. L'opérateur gradient est borné, de norme $L \leq 2\sqrt{2}$, avec L tendant vers cette borne supérieure lorsque N_x ou N_y tendent vers $+\infty$.

Les trois jeux de paramètres sont donc donnés par

1. ADMM classique : $\lambda = \lambda' = 4\sqrt{\mu}$;
2. ADMM modifié :

$$(\lambda, \lambda') = \left(\frac{1}{2} \left(\sqrt{1 + \frac{32}{\mu}} + 1 \right), \frac{1}{2} \left(\sqrt{1 + \frac{32}{\mu}} - 1 \right) \right).$$

Mesure de la convergence Pour mesurer la convergence, on utilise cette fois deux outils distincts : la différence entre l'énergie primale et son minimum et la distance quadratique en v_n^h et le minimiseur v^* du problème. La première mesure dépendant de la valeur de la solution v^* , cette dernière est obtenue expérimentalement en faisant tourner sur un temps long la méthode PDHG sur-relaxé.

Résultats expérimentaux La figure 5.4 présente l'image initiale bruitée u^h ainsi que le débruitage v^* réalisé en résolvant le problème (5.50). On a choisi $\mu = 2$ pour cette expérience. On compare ensuite les performances de l'ADMM classique, l'ADMM modifié, le PDHG sur-relaxé et FISTA. La figure 5.5 présente quant à elle la convergence de l'énergie et celle de la variable primale pour chaque expérience. Dans chaque cas, on affiche également la courbe de décroissance théorique, qui, en échelle logarithmique, est une droite de pente $\ln(\tilde{\omega})$ pour la distance au minimiseur.

5.4.5 Discussion

ADMM classique versus ADMM modifié Les deux expériences présentées aux paragraphes précédents montrent dans les deux cas que la version modifiée de l'ADMM proposée dans ce chapitre conduit à de meilleures performances que l'ADMM classique. Pour une comparaison juste, on a choisi pour l'ADMM classique et une des versions de l'ADMM modifié les paramètres conduisant théoriquement au meilleur taux de convergence. On observe alors que, tant pour l'énergie primale et pour le *primal-dual gap*, l'introduction d'un paramètre $\lambda' \neq \lambda$ permet effectivement d'accélérer la décroissance.

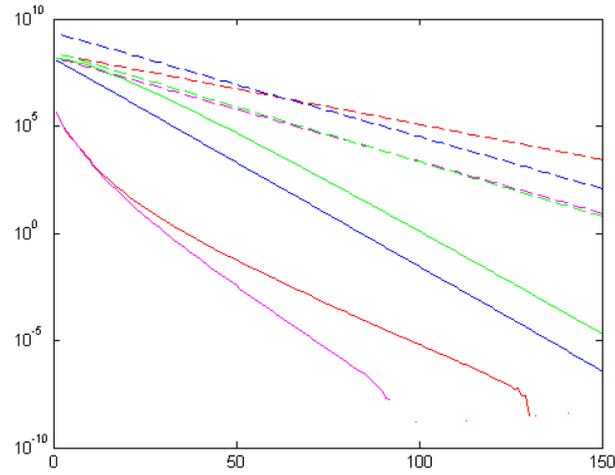
Comparaison avec les taux théoriques Dans le premier exemple du problème quadratique, les performances observées empiriquement sont bien meilleures que celles attendues (qui sont affichées en pointillés). Cela peut expliquer par le fait que le problème est beaucoup plus régulier que ceux concernés par les différentes preuves de convergence démontrées dans ce chapitre.

Dans la régularisation TV-HUBER, on observe pour l'ADMM, l'ADMM modifié et PDHG une bonne corrélation entre les courbes théoriques et les courbes empiriques dans le cas de la convergence de la variable primale (même si cette convergence n'a pas été formellement démontrée dans le cas des ADMM). En revanche, il apparaît que la convergence de l'énergie et du *gap* soit meilleure en pratique qu'en théorie, puisqu'elle semble afficher le même taux que celle de la variable primale.

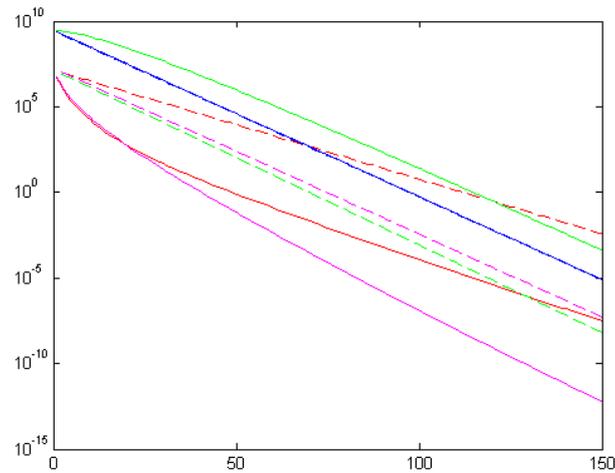
Oscillations de FISTA et de PDHG accéléré Pour FISTA et PDHG sur-relaxé, on observe des oscillations dans le problème quadratique. Dans le cas de FISTA, ce phénomène a déjà été observé [12] pour des problèmes similaires. Ces oscillations sont dues à l'étape de sur-relaxation, et apparaissent lorsque θ est choisi trop grand en comparaison des valeurs propres de l'opérateur $mI_N + (M - m)K_N^*K_N$.

Dans le cas de l'algorithme primal-dual, de telles oscillations n'ont pas été rapportées à notre connaissance. Il est vraisemblable qu'elles soient également dues à la sur-relaxation, pour des raisons similaires à celles intervenant dans FISTA. Cependant, à l'heure où ce manuscrit est rédigé, l'étude théorique de ce phénomène n'a pas encore abouti.

Mis à part l'apparition d'oscillations dans le cas quadratique mal conditionné, qui ralentissent localement la convergence des algorithmes PDHG et FISTA, il doit être noté que ces deux algorithmes ne font pas intervenir l'inversion d'une matrice (qui, dans le cas du débruitage TV-HUBER, peut s'avérer être d'une taille très importante). Ainsi, une itération dans ces deux algorithmes est bien moins coûteuse en temps de calculs. En d'autres termes, même si la convergence en termes de nombre d'itérations peut être meilleure pour l'ADMM modifié, celle est très inférieure lorsqu'elle est mesurée en temps total de calculs.



(a) Énergie $E(x_n) - E(x^*)$



(b) Distance au minimiseur $\|x_n - x^*\|^2$

FIGURE 5.5 – Mesure de la convergence pour l'ADMM (rouge), l'ADMM modifié (magenta), PDHG sur-relaxé (bleu) et FISTA (vert) pour le débruitage TV-HUBER. En trait plein, les mesures empiriques, en pointillé, les valeurs théoriques (décroissance linéaire en ω pour l'énergie et en $\tilde{\omega}$ pour la convergence de la variable primale). Les ordonnées sont en échelle logarithmique.

Conclusion

On a proposé dans ce chapitre une version modifiée de l'algorithme ADMM en relaxant légèrement le choix du paramètre de la version classique de l'algorithme, en introduisant un second paramètre. Dans le cas fortement convexe, on a démontré, en écrivant les itérations de l'ADMM sous la forme d'un algorithme primal-dual connu sous le nom de PDHG, la convergence de ce nouvel algorithme, sous certaines conditions sur les deux paramètres.

L'introduction de ce second paramètre, s'il est correctement choisi, conduit à une meilleure convergence théorique que l'ADMM classique. On a déterminé pour ces deux versions les paramètres conduisant au meilleur taux de convergence théorique, ainsi que, pour la version modifiée, des paramètres associés à un taux sous-optimal mais permettant de conserver un contrôle sur la convergence des deux variables.

Des tests expérimentaux, dont l'un sur le débruitage, confirment qu'en pratique, la version modifiée de l'ADMM présente de meilleures performances que la version classique. Dans un cas simple mais mal-conditionné, elle affiche par ailleurs un meilleur comportement que deux autres algorithmes accélérés classiques, qui sont l'algorithme primal-dual PDHG sur-relaxé et FISTA (accéléré dans le cas fortement convexe). Ces deux derniers présentent en effet des oscillations qui dégradent leur vitesse de convergence. Néanmoins, dans des cas moins pathologiques, même si les deux versions d'ADMM étudiées offrent des convergences comparables à ces deux algorithmes en termes de nombre d'itérations, l'inversion d'un opérateur nécessaire pour l'ADMM le rend moins compétitif lorsque la convergence est mesurée en termes de temps de calcul.

Références

- [1] Hedy ATTOUCH and Mohamed SOUEYCATT. Augmented lagrangian and proximal alternating direction methods of multipliers in Hilbert spaces. applications to games, PDE's and control. *Pacific Journal of Optimization*, 5(1) :17–37, 2008.
- [2] Amir BECK and Marc TEBoulLE. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1) :183–202, 2009.
- [3] Stephen BOYD, Neal PARIKH, Eric CHU, Borja PELEATO, and Jonathan ECKSTEIN. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1) :1–122, 2011.
- [4] Antonin CHAMBOLLE and Thomas POCK. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1) :120–145, 2011.
- [5] Antonin CHAMBOLLE and Thomas POCK. On the ergodic convergence rates of a first-order primal-dual algorithm. *Mathematical Programming*, pages 1–35, 2015.
- [6] Antonin CHAMBOLLE and Thomas POCK. An introduction to continuous optimization for imaging. *Acta Numerica*, 25 :161–319, 5 2016.

-
- [7] Patrick L. COMBETTES and Jean-Christophe PESQUET. Proximal splitting methods in signal processing. In *Fixed-point algorithms for inverse problems in science and engineering*, pages 185–212. Springer, 2011.
- [8] Ernie ESSER, Xiaoqun ZHANG, and Tony CHAN. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4) :1015–1046, 2010.
- [9] Daniel GABAY and Bertrand MERCIER. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1) :17–40, 1976.
- [10] Roland GLOWINSKI and A. MARROCCO. Sur l’approximation, par éléments finis d’ordre un, et la résolution, par pénalisation-dualité d’une classe de problèmes de dirichlet non linéaires. *Revue Française d’Automatique, Informatique, Recherche Opérationnelle. Analyse Numérique*, 9(2) :41–76, 1975.
- [11] Yurii NESTEROV. *Introductory lectures on convex optimization : A basic course*, volume 87. Springer Science & Business Media, 2013.
- [12] Brendan O’DONOGHUE and Emmanuel CANDES. Adaptive restart for accelerated gradient schemes. *Foundations of computational mathematics*, 15(3) :715–732, 2015.
- [13] Thomas POCK, Daniel CREMERS, Horst BISCHOF, and Antonin CHAMBOLLE. An algorithm for minimizing the MUMFORD-SHAH functional. In *IEEE International Conference on Computer Vision*, pages 1133–1140. IEEE, 2009.
- [14] Mingqiang ZHU and Tony CHAN. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA CAM Report*, pages 08–34, 2008.

