Stratégies multicouche, avec mémoire, et à métrique variable en méthodes de point fixe pour l'éclatement d'opérateurs monotones et l'optimisation

Lilian Glaudin

SORBONNE UNIVERSITÉ

THÈSE DE DOCTORAT

Spécialité : Mathématiques

Laboratoire Jacques-Louis Lions | UMR 7598

Stratégies multicouche, avec mémoire, et à métrique variable en méthodes de point fixe pour l'éclatement d'opérateurs monotones et l'optimisation

LILIAN GLAUDIN

soutenue le 20 février 2019 devant le jury composé de :

Samir ADLY	*Examinateur*
Radu Ioan BOŢ	*Rapporteur*
Antonin CHAMBOLLE	*Examinateur*
Patrick L. COMBETTES	*Directeur de thèse*
Charles DOSSAL	*Rapporteur*
Sylvain SORIN	*Président*

Remerciements

Cette thèse fut un travail de longue haleine qui n'aurait pas été possible sans l'aide et le soutien d'un bon nombre de personnes. En premier lieu, je remercie Patrick Combettes qui a accepté de diriger ma thèse. Il a beaucoup influencé ma façon d'appréhender les mathématiques, me nourrissant de ses connaissances et de sa rigueur. Malgré la distance qui nous a séparés durant deux ans, il a su être se montrer présent.

Je souhaite remercier les membres de mon jury de thèse, en commençant par Radu Ioan Boţ et Charles Dossal qui ont accepté de rapporter ma thèse. Je remercie également Samir Adly, Antonin Chambolle et Sylvain Sorin d'avoir accepté de faire partie de mon jury de thèse.

Le Laboratoire Jacques-Louis Lions, qui fut mon lieu de travail pendant la majorité de ces trois ans, n'est pas vraiment un repère d'optimiseurs. Pourtant, cela ne m'a pas empêché d'y faire des rencontres d'un grand soutien, idéales pour préparer un doctorat. Je remercie évidemment tous les doctorants, passés et présents, de m'avoir supporté, et particulièrement ceux qui se sont risqués aux réflexions mathématiques loin de leur domaine de prédilection. Merci à Geneviève, Anouk, Nicolas, Jules, Valentin, Christophe et tous les autres. Je remercie aussi le secrétariat et l'équipe informatique, toujours efficaces et sympathiques.

Afin de me rapprocher de mon directeur, je suis parti pour les États-Unis. Ce fut l'occasion de rencontrer des gens formidables tant sur le plan intellectuel qu'amical. Je souhaite remercier tout particulièrement Farid, Zev et Georgy pour toutes ces discussions mathématiques et leur accueil chaleureux.

Il n'est pas toujours facile de relire son propre travail. Je remercie donc tous ceux qui m'ont soutenu durant la rédaction et qui ont relu mon manuscrit, parfois sans y comprendre un mot.

Pour finir, je remercie ma famille de m'avoir soutenu dans les moments difficiles, et Marne bien sur, sans qui rien de tout ça n'aurait été possible.

Je dédie cette thèse à ma maman, partie trop tôt.

Paris, le 20 février 2019

Table des matières

Résumé

Stratégies multicouche, avec mémoire, et à métrique variable en méthodes de point fixe pour l'éclatement d'opérateurs monotones et l'optimisation

Plusieurs stratégies sans liens apparents coexistent pour mettre en œuvre les algorithmes de résolution de problèmes d'inclusion monotone dans les espaces hilbertiens. Nous proposons un cadre synthétique permettant d'englober diverses approches algorithmiques pour la construction de point fixe, clarifions et généralisons leur théorie asymptotique, et concevons de nouveaux schémas itératifs pour l'analyse non linéaire et l'optimisation convexe. Notre méthodologie, qui est ancrée sur un modèle de compositions de quasicontractions moyennées, nous permet de faire avancer sur plusieurs fronts la théorie des algorithmes de point fixe et d'impacter leurs domaines d'applications. Des exemples numériques sont fournis dans le contexte de la restauration d'image, où nous proposons un nouveau point de vue pour la formulation des problèmes variationnels.

Mots-clés : algorithme inertiel, algorithme à métrique variable, algorithme multicouche, algorithme avec mémoire, algorithme de point fixe, éclatement d'opérateur monotone, opérateur moyenné, optimisation convexe, traitement de l'image.

Abstract

Strategies with multiple layers, memory, and variable metric in fixed point methods for monotone operator splitting and optimization

Several apparently unrelated strategies coexist to implement algorithms for solving monotone inclusions in Hilbert spaces. We propose a synthetic framework for fixed point construction which makes it possible to capture various algorithmic approaches, clarify and generalize their asymptotic behavior, and design new iterative schemes for nonlinear analysis and convex optimization. Our methodology, which is anchored on an averaged quasinonexpansive operator composition model, allows us to advance the theory of fixed point algorithms on several fronts, and to impact their application fields. Numerical examples are provided in the context of image restoration, where we propose a new viewpoint on the formulation of variational problems.

Key words : algorithm with memory, averaged operator, convex optimization, fixed point algorithm, image processing, inertial algorithm, monotone operator splitting, multilayer algorithm, variable metric algorithm.

Notations et glossaire

Les notations suivantes sont utilisées dans le manuscrit.

Notations générales

- $\mathcal{H}, \mathcal{G}$: Espaces hilbertiens réels.
- $\langle \cdot \mid \cdot \rangle$: Produit scalaire d'un espace hilbertien.
- $\| \cdot \|$: Norme d'un espace hilbertien.
- $B(x;\rho)$: Boule de centre $x \in \mathcal{H}$ et de rayon $\rho \in \,]0,+\infty[$.
- Id : Opérateur identité.
- $2^{\mathcal{H}}$: Ensemble des parties de $\mathcal{H}$.
- $\Gamma_0(\mathcal{H})$: Ensemble des fonctions de $\mathcal{H}$ dans $]-\infty,+\infty]$, convexes, propres, et semi-continues inférieurement.
- $\mathcal{B}\,(\mathcal{H},\mathcal{G})$: Espace des opérateurs linéaires bornés de $\mathcal{H}$ dans $\mathcal{G}$.
- $\mathcal{B}\,(\mathcal{H}) = \mathcal{B}\,(\mathcal{H},\mathcal{H})$.
- L^* : Adjoint d'un opérateur $L \in \mathcal{B}\,(\mathcal{H},\mathcal{G})$.
- $\mathcal{S}\,(\mathcal{H}) = \big\{L \in \mathcal{B}\,(\mathcal{H}) \;\big|\; L^* = L\big\}$.
- $\to$: Convergence forte.
- $\rightharpoonup$: Convergence faible.
- $\ell_+(\mathbb{N}) = \Big\{(\xi_n)_{n\in\mathbb{N}} \in [0,+\infty[^{\mathbb{N}}\Big\}$.
- $\ell^\infty(\mathbb{N}) = \Big\{(\xi_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\Big|\; \sup_{n\in\mathbb{N}} |\xi_n| < +\infty\Big\}$.
- $\ell^1(\mathbb{N}) = \Big\{(\xi_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\Big|\; \sum_{n\in\mathbb{N}} |\xi_n| < +\infty\Big\}$.
- $\ell^2(\mathbb{N}) = \Big\{(\xi_n)_{n\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \;\Big|\; \sum_{n\in\mathbb{N}} |\xi_n|^2 < +\infty\Big\}$.
- $[\,\cdot\,]^+$: Partie positive d'un réel.

Notations relatives à un sous-ensemble C de $\mathcal{H}$

- ι_C : Fonction indicatrice de C :

$$\iota_C\colon \mathcal{H} \to [-\infty,+\infty] : x \mapsto \begin{cases} 0, & \text{si } \; x \in C; \\ +\infty, & \text{si } \; x \notin C. \end{cases}$$

- $\operatorname{int} C$: Intérieur de C.
- $\overline{C}$: Adhérence de C.
- $d_C\colon \mathcal{H} \to [0,+\infty]\colon x \mapsto \inf_{x\in\mathcal{H}} d(x,C)$: Fonction distance à C.
- Supposons C convexe, fermé, non vide. $P_C = \operatorname{proj}_C\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \operatorname{argmin}_{p\in C} \|x-p\|$: Opérateur de projection sur C.

Notations relatives à un opérateur $T\colon \mathcal{H} \to \mathcal{H}$

- $\operatorname{Fix} T = \{x \in \mathcal{H} \mid Tx = x\}$: Points fixes de T.
- T est une contraction si
$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\| \leqslant \|x - y\|.$$
- T est une quasicontraction si
$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\| \leqslant \|x - y\|.$$
- Soit $\alpha \in \left]0,1\right]$. T est α-moyenné si
$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\|^2 \leqslant \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|(\operatorname{Id} - T)x - (\operatorname{Id} - T)y\|^2.$$
- Soit $\alpha \in \left]0,1\right]$. T est α-quasimoyenné si
$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\|^2 \leqslant \|x - y\|^2 - \frac{1-\alpha}{\alpha}\|Tx - x\|^2.$$
- Soit $\beta \in \left]0,+\infty\right[$. T est β-cocoercif si
$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Tx - Ty\rangle \geqslant \beta\|Tx - Ty\|^2.$$
- Soit $u \in \mathcal{H}$. T est demi-fermé en u si pour toute suite $(x_n)_{n\in\mathbb{N}}$ dans $\mathcal{H}$ et pour tout $x \in \mathcal{H}$ tels que $x_n \rightharpoonup x$ et $Tx_n \to u$, on a $Tx = u$. De plus T est demi-fermé s'il est demi-fermé en tout point de $\mathcal{H}$.

Notations relatives à un opérateur multivoque $M\colon \mathcal{H} \to 2^{\mathcal{H}}$

- $\operatorname{dom} M = \{x \in \mathcal{H} \mid Mx \neq \varnothing\}$: Domaine de M.
- $\operatorname{gra} M = \{(x,u) \in \mathcal{H} \times \mathcal{H} \mid u \in Mx\}$: Graphe de M.
- $M^{-1}\colon \mathcal{H} \to 2^{\mathcal{H}}\colon u \mapsto \{x \in \mathcal{H} \mid u \in Mx\}$: Inverse de M.
- $\operatorname{zer} M = \{x \in \mathcal{H} \mid 0 \in Mx\}$: Zéros de M.

- ran $M = \{u \in \mathcal{H} \mid (\exists x \in \mathcal{H})\ \ u \in Mx\}$: Image de M.
- $s\colon \operatorname{dom} M \to \mathcal{H}$ tel que $(\forall x \in \operatorname{dom} M)\ s(x) \in Mx$: Sélection de M.
- M est monotone si

$$(\forall (x_1,u_1) \in \operatorname{gra} M)(\forall (x_2,u_2) \in \operatorname{gra} M) \quad \langle x_1 - x_2 \mid u_1 - u_2\rangle \geqslant 0.$$

- M est maximalement monotone si M est monotone et s'il n'existe pas d'opérateur monotone de $\mathcal{H}$ dans $2^{\mathcal{H}}$ dont le graphe contient strictement gra M.
- $J_M = (\operatorname{Id} + M)^{-1}$: Résolvante de M.
- Soit $x \in \mathcal{H}$. M est demi-régulier en x si pour toute suite $((x_n,u_n))_{n\in\mathbb{N}}$ dans gra M et pour tout $u \in Mx$ tels que $x_n \rightharpoonup x$ et $u_n \to u$, on a $x_n \to x$.

Notations relatives à une fonction $f \in \Gamma_0(\mathcal{H})$

- dom $f = \{x \in \mathcal{H} \mid f(x) < +\infty\}$: Domaine de f.
- Argmin f : Ensemble des minimiseurs de f sur $\mathcal{H}$.
- $f^*\colon \mathcal{H} \to \left]-\infty,+\infty\right] : u \mapsto \sup\limits_{x\in\mathcal{H}} \big(\langle x \mid u\rangle - f(x)\big)$: Conjuguée de f.
- $\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid u\rangle + f(x) \leqslant f(y)\}$: Sous-différentiel de f.
- Soit $\varepsilon \in \left]0,+\infty\right[$. $\partial_\varepsilon f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \{u \in \mathcal{H} \mid (\forall y \in \mathcal{H})\ \langle y - x \mid u\rangle + f(x) \leqslant f(y) + \varepsilon\}$: ε-sous-différentiel de f.
- f est lisse si dom $f = \mathcal{H}$, si f est différentiable et si ∇f est lipschitzien.
- Soit C un sous-ensemble C de $\mathcal{H}$. f est uniformément quasiconvexe sur C si il existe une fonction $\phi\colon\ \left[0,+\infty\right[\to [0,+\infty]$ croissante ne s'annulant qu'en 0 et si

$$(\forall x \in C)(\forall y \in C)(\forall \alpha \in \left]0,1\right[) \qquad f(\alpha x + (1-\alpha)y) + \alpha(1-\alpha)\phi(\|x-y\|) \leqslant \max\{f(x), f(y)\}. \quad (1)$$

Chapitre 1

Introduction

1.1 Méthodes de point fixe et éclatement d'opérateurs

Dans de nombreux domaines de l'analyse non linéaire, les méthodes de point fixe sont un moyen naturel de modéliser, d'analyser et de résoudre des problèmes complexes [7, 12, 29, 45]. Dans [14] il est montré que des méthodes classiques d'optimisation et de résolution d'inclusions monotones peuvent être modélisées et étudiées en tant que méthodes de point fixe. Ces dernières interviennent naturellement dans de nombreux champs des mathématiques appliquées : statistiques, apprentissage automatique, traitement du signal, équations aux dérivées partielles, problèmes inverses, transport optimal, optimisation [2, 4, 15, 19, 20, 23, 27, 28, 30, 34, 39, 42]. La notion d'opérateur moyenné, introduite dans [5], joue un rôle fondamental dans notre travail.

Définition 1.1 Soient $\mathcal{H}$ un espace de Hilbert réel, $\alpha \in \,]0,1]$ et $T\colon \mathcal{H} \to \mathcal{H}$. Alors T est α-moyenné s'il existe une contraction $R\colon \mathcal{H} \to \mathcal{H}$ telle que $T = (1-\alpha)\,\mathrm{Id} + \alpha R$.

Un problème générique de point fixe est de

$$\text{trouver } \; x \in \mathcal{H} \; \text{ tel que } \; x = Tx, \quad \text{où } \; T \; \text{ est une contraction.} \tag{1.1}$$

Pour plus de flexibilité, il est préférable de considérer une formulation multicouche où l'on décompose l'opérateur T en opérateurs plus simples. En particulier, ceci permet de modéliser de manière plus souple les problèmes d'éclatement pour les inclusions monotones. Pour tout $i \in \{1, \ldots, m\}$, soit $T_i\colon \mathcal{H} \to \mathcal{H}$ un opérateur moyenné. L'objectif est de

$$\text{trouver } \; x \in \mathcal{H} \; \text{ tel que } \; x = (T_1 \circ \cdots \circ T_m)x. \tag{1.2}$$

Plus généralement, nous allons considérer une famille de problèmes de la forme (1.2) et obtenir la formulation suivante déjà proposée dans [14, 24].

Problème 1.2 Soit $m \geqslant 2$ un entier. Pour tout $i \in \{1, \dots, m\}$ et tout $n \in \mathbb{N}$, soit $\alpha_{i,n} \in]0,1[$ et soit $T_{i,n} : \mathcal{H} \to \mathcal{H}$ un opérateur $\alpha_{i,n}$-moyenné. L'objectif est de

$$\text{trouver} \quad x \in \bigcap_{n \in \mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad \text{où} \quad (\forall n \in \mathbb{N}) \quad T_n = T_{1,n} \circ \cdots \circ T_{m,n}. \tag{1.3}$$

Pour résoudre ce problème, le théorème suivant permet des relaxations modélisées par la suite $(\lambda_n)_{n\in\mathbb{N}}$. Des erreurs lors du calcul de chaque opérateur y sont également tolérées.

Théorème 1.3 *[24, Theorem 3.5(iii)] On considère le cadre du Problème 1.2. Soit $\varepsilon \in]0, 1/2[$. Pour tout $n \in \mathbb{N}$, soit $\phi_n \in]0,1[$ une constante de moyennage de T_n, soit $\lambda_n \in]0, (1-\varepsilon)(1+\varepsilon\phi_n)/\phi_n]$ et, pour tout $i \in \{1, \dots, m\}$, soit $e_{i,n} \in \mathcal{H}$. Pour tout $i \in \{1, \dots, m\}$, on suppose que $\sum_{n\in\mathbb{N}} \lambda_n \|e_{i,n}\| < +\infty$. Pour tout $n \in \mathbb{N}$, on itère*

$$x_{n+1} = x_n + \lambda_n\Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}x_n + e_{m,n}) + e_{m-1,n}\cdots\big) + e_{2,n}\Big) + e_{1,n} - x_n\Big). \tag{1.4}$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n$ si et seulement si tout point d'accumulation séquentiel faible de $(x_n)_{n\in\mathbb{N}}$ est dans $\bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n$.

En théorie des opérateurs monotones, la construction d'un zéro d'une somme d'opérateurs est un problème fondamental [8, 11, 31, 41, 43, 46, 47], qui intervient notamment dans la résolution des problèmes de minimisation (voir [16] et sa bibliographie). Le résultat suivant est un résultat de base dans ce domaine.

Théorème 1.4 *[41, Theorem 1] Soit $M : \mathcal{H} \to 2^{\mathcal{H}}$ un opérateur maximalement monotone tel que $\operatorname{zer} M \neq \varnothing$. Soient $\gamma \in]0, +\infty[$ et $x_0 \in \mathcal{H}$. On itère*

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma M} x_n. \tag{1.5}$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un zéro de M.

Une formulation générique à deux opérateurs est la suivante.

Problème 1.5 Soient $A : \mathcal{H} \to 2^{\mathcal{H}}$ et $B : \mathcal{H} \to 2^{\mathcal{H}}$ deux opérateurs maximalement monotones. Supposons que $\operatorname{zer}(A + B) \neq \varnothing$. L'objectif est de

$$\text{trouver} \quad x \in \mathcal{H} \quad \text{tel que} \quad 0 \in Ax + Bx. \tag{1.6}$$

En principe, si $A + B$ est maximalement monotone, on peut poser $M = A + B$ et résoudre le Problème 1.5 via le Théorème 1.4. Cependant cette approche n'est généralement pas possible car le calcul de la résolvante de M est trop complexe. On utilise une méthode dite d'éclatement, dans laquelle les opérateurs A et B sont activés de manière indépendante. L'algorithme de Douglas-Rachford permet de résoudre le Problème 1.5 dans ce cadre [14, 18, 26, 31].

Théorème 1.6 *[8, Theorem 26.11] On considère le cadre du Problème 1.5. Soient $\gamma \in]0,+\infty[$ et $y_0 \in \mathcal{H}$. On itère*

$$
\begin{array}{l}
\text{pour } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
x_n = J_{\gamma B} y_n \\
z_n = J_{\gamma A}(2x_n - y_n) \\
y_{n+1} = y_n + z_n - x_n.
\end{array}
\right.
\end{array}
\tag{1.7}
$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un zéro de $A + B$.

Dans certaines situations, l'opérateur B est cocoercif, on peut alors utiliser la méthode explicite-implicite ci-dessous [14, 21, 23, 31].

Théorème 1.7 *[14, Corollary 6.5] On considère le cadre du Problème 1.5 et on suppose que B est cocoercif. Soient $\beta \in]0,+\infty[$ la constante de cocoercivité de B, $(\gamma_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0, 2\beta[$ telle que $0 < \underline{\lim}\, \gamma_n \leqslant \overline{\lim}\, \gamma_n < 2\beta$, et $x_0 \in \mathcal{H}$. On itère*

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = J_{\gamma_n A}(x_n - \gamma_n B x_n). \tag{1.8}
$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un zéro de $A + B$.

On peut résoudre les problèmes d'éclatement d'opérateurs monotones par le formalisme des opérateurs moyennés [14, 24]. Considérons donc une version à deux couches du Problème 1.2.

Algorithme 1.8 Pour tout $n \in \mathbb{N}$, soient $T_{1,n} \colon \mathcal{H} \to \mathcal{H}$ et $T_{2,n} \colon \mathcal{H} \to \mathcal{H}$ deux opérateurs moyennés et soit $x_0 \in \mathcal{H}$. On itère

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = (T_{1,n} \circ T_{2,n}) x_n. \tag{1.9}
$$

Si, pour tout $n \in \mathbb{N}$, on prend $T_{1,n} = J_{\gamma_n A}$ et $T_{2,n} = \mathrm{Id} - \gamma_n B$, alors (1.9) se ramène à la méthode explicite-implicite (1.8). On voit ici l'intérêt des méthodes de point fixe multicouches pour analyser et résoudre les problèmes d'éclatement d'opérateurs monotones.

Dans les algorithmes ci-dessus, à l'itération $n \in \mathbb{N}$, x_{n+1} est calculé en utilisant exclusivement l'itéré courant x_n. Il est aussi possible de construire des algorithmes avec mémoire, dans lesquels x_{n+1} est calculé à partir de x_0, x_1, $\ldots$, x_n. C'est le cas notamment de la théorie des moyennes de Mann [6, 17, 25, 32, 33, 38]. Pour étudier ces algorithmes, la notion de matrice concentrante introduite dans [17] est pertinente.

Définition 1.9 [17, Definition 2.1] Soit $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ une matrice de réels positifs telle que

$$
\lim_{n\to+\infty} \sum_{j=0}^{n} \mu_{n,j} = 1 \quad \text{et} \quad (\forall j \in \mathbb{N}) \quad \lim_{n\to+\infty} \mu_{n,j} = 0. \tag{1.10}
$$

Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est *concentrante* si toute suite $(\xi_n)_{n\in\mathbb{N}}$ à valeurs dans $[0,+\infty[$ qui satisfait

$$\left(\exists\,(\varepsilon_n)_{n\in\mathbb{N}}\in[0,+\infty[^{\mathbb{N}}\right)\ \sum_{n\in\mathbb{N}}\varepsilon_n<+\infty \quad \text{et} \quad (\forall n\in\mathbb{N})\quad \xi_{n+1}\leqslant\sum_{j=0}^{n}\mu_{n,j}\xi_j+\varepsilon_n \tag{1.11}$$

converge.

Ce cadre permet de trouver un point fixe commun à une famille d'opérateurs $(T_n)_{n\in\mathbb{N}}$ comme suit.

Théorème 1.10 *[17, Theorem 3.5(i)] Soit $\varepsilon\in\,]0,1[$, soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice concentrante et, pour tout $n\in\mathbb{N}$, soit $T_n\colon\mathcal{H}\to\mathcal{H}$ un opérateur $1/2$-quasimoyenné et soit $\lambda_n\in\,]\varepsilon,2-\varepsilon[$. Supposons que $\bigcap_{n\in\mathbb{N}}\operatorname{Fix}T_n\neq\varnothing$ et soit $x_0\in\mathcal{H}$. On itère*

$$\begin{array}{l}\text{pour } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}\overline{x}_n=\displaystyle\sum_{j=0}^{n}\mu_{n,j}x_j\\ x_{n+1}=\overline{x}_n+\lambda_n\left(T_n\overline{x}_n-\overline{x}_n\right).\end{array}\right.\end{array} \tag{1.12}$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\bigcap_{n\in\mathbb{N}}\operatorname{Fix}T_n$ si et seulement si tout point d'accumulation séquentiel faible de $(x_n)_{n\in\mathbb{N}}$ est dans $\bigcap_{n\in\mathbb{N}}\operatorname{Fix}T_n$.

Ce type de méthode a coexisté avec un autre type d'approche avec mémoire, à savoir les algorithmes dits inertiels qui utilisent un poids négatif sur un itéré passé [1, 3, 9, 10, 13, 35, 37, 40]. La méthode explicite-implicite (1.8) a ainsi été étudiée avec un terme inertiel, dans le cas particulier de la minimisation de la somme de deux fonctions convexes f et g, où g est lisse. Ceci correspond à poser $A=\partial f$ et $B=\nabla g$ dans le Problème 1.5.

Algorithme 1.11 (Algorithme explicite-implicite inertiel) Soient $\beta\in\,]0,+\infty[$, $\gamma\in\,]0,\beta]$ et $(\eta_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,1[$. Soit $f\in\Gamma_0(\mathcal{H})$, soit $g\colon\mathcal{H}\to\mathbb{R}$ une fonction convexe lisse de gradient $1/\beta$-lipschitzien telle que $\operatorname{Argmin}(f+g)\neq\varnothing$, et soit $x_{-1}=x_0\in\mathcal{H}$. On itère

$$\begin{array}{l}\text{pour } n=0,1,\ldots\\ \left\lfloor\begin{array}{l}y_n=x_n+\eta_n(x_n-x_{n-1})\\ x_{n+1}=\operatorname{prox}_{\gamma f}\big(y_n-\gamma\nabla g(y_n)\big).\end{array}\right.\end{array} \tag{1.13}$$

À l'itération $n\in\mathbb{N}$, au lieu d'appliquer les opérateurs directement à x_n, on les applique à y_n, que nous interprétons comme une combinaison affine entre x_n et x_{n-1}. Remarquons qu'on impose ici $\gamma\leqslant\beta$ contrairement au Théorème 1.7 où $\gamma<2\beta$. Pour certaines suites $(\eta_n)_{n\in\mathbb{N}}$, cet algorithme produit une suite $((f+g)(x_n)-\min(f+g)(\mathcal{H}))_{n\in\mathbb{N}}$

qui réalise un taux optimal au sens de [36]. De plus, si pour tout $n \in \mathbb{N}$, $\eta_n = (n-1)/(n+3)$, la convergence faible des itérés vers un minimiseur est garantie [13].

Un de nos principaux objectifs va être de proposer un cadre général qui unifie et étend ces techniques et leurs champs d'application.

1.2 Motivations et organisation de la thèse

De nombreuses stratégies sans liens apparents coexistent pour résoudre les problèmes d'inclusion monotone. Dans le but de simplifier et rationaliser l'état de l'art autour de principes élémentaires, nous proposons une théorie cohérente capable d'englober diverses techniques algorithmiques dans un cadre synthétique. Par exemple, les méthodes inertielles et celles des moyennes de Mann sont toujours étudiées de manière distincte. De même, les techniques d'itérations à métriques variables n'ont été étudiées que dans le cadre d'algorithmes spécifiques. Nous développons des outils pour analyser ces différentes constructions algorithmiques conjointement, clarifions et généralisons la théorie asymptotique de ces méthodes, et proposons de nouvelles méthodes itératives pour l'analyse non linéaire et l'optimisation. Notre méthodologie, ancrée sur un modèle de compositions de quasicontractions moyennées, nous permet de faire avancer simultanément sur plusieurs fronts la théorie des algorithmes de point fixe et de leurs applications aux inclusions monotones. Des exemples numériques sont proposés dans le cadre de la restauration d'image, où nous exposons un nouveau point de vue pour la formulation des problèmes variationnels.

La thèse est organisée comme suit.

Dans le Chapitre 2, nous nous intéressons aux algorithmes dont les règles de mise à jour utilisent l'ensemble des itérés déjà générés. Nous présentons un cadre général d'itération de point fixe dont le but est d'englober d'une part les méthodes de point fixe générales et, d'autre part, deux méthodes avec mémoire qui semblaient indépendantes, à savoir les moyennes de Mann et les techniques inertielles. Nous fournissons de nombreux exemples d'applications de notre théorème principal, et montrons qu'il couvre une large part des résultats connus dans ces divers domaines. Nous revisitons alors l'algorithme explicite-implicite, l'algorithme de Peaceman-Rachford, mais aussi par exemple l'algorithme du sous-gradient projeté de Polyak.

L'objectif du Chapitre 3 est de compléter l'analyse du Chapitre 2 en étudiant plus en détail la notion de matrice concentrante. Nous en proposons une spécialisation, appelée *superconcentrante*, qui nous permet d'obtenir des résultats de sommabilité. Nous fournissons des exemples de telles matrices et nous proposons aussi de nouvelles matrices inertielles. Finalement, nous montrons l'intérêt d'obtenir la sommabilité en proposant des exemples d'algorithmes avec mémoire dont on ne sait pas prouver la convergence sans matrices superconcentrantes, comme l'algorithme de Krasnosel'skiĭ-Mann, ou en l'utilisant pour garantir la convergence forte d'une suite.

Le Chapitre 4 présente une extension des méthodes de point fixe au cadre où la métrique sous-jacente change à chaque itération. De telles variantes ont notamment été étudiées dans le cas de l'algorithme explicite-implicite [22, 44]. Nous appliquons ces techniques pour proposer un algorithme multicouche sur-relaxé d'opérateurs moyennés avec métrique variable. Nous présentons plusieurs nouvelles méthodes et, en particulier, une application à l'algorithme explicite-implicite sur-relaxé à métrique variable, et à un algorithme explicite-implicite sur-relaxé dans l'espace primal-dual pour résoudre un problème composite.

Enfin, le Chapitre 5 est une réflexion originale sur le choix des algorithmes dans les applications. Sur des exemples simples, nous avons constaté l'efficacité de l'algorithme de Douglas-Rachford (1.7) par rapport à l'algorithme explicite-implicite (1.8), pourtant habituellement privilégié sur ce type d'exemple. Ce constat nous conduit à remettre en cause la stratégie qui consiste à utiliser systématiquement des pas explicites sur les termes lisses dans les méthodes d'éclatement en optimisation convexe du premier ordre. Cette proposition peut surprendre car l'utilisation explicite d'opérateurs monotones se fait presque systématiquement en raison des propriétés supplémentaires (régularité, cocoercivité) dont ils jouissent. Nous fournissons un grand nombre de fonctions lisses utiles dans les applications dont on sait calculer l'opérateur proximal. Plusieurs simulations numériques en traitement de l'image viennent illustrer cette étude, montrant qu'il peut être avantageux d'utiliser des algorithmes entièrement proximaux, donc implicites, plutôt que des méthodes qui font intervenir des pas explicites.

1.3 Contributions principales

Les contributions principales de la thèse sont les suivantes.

- Conception et étude de la première méthode de point fixe multicouche avec mémoire sur-relaxée avec erreurs. (Chapitre 2)
- Extension des algorithmes du cadre des opérateurs moyennés au cadre des opérateurs quasimoyennés et contractants. (Chapitre 2)
- Généralisation de nombreux algorithmes d'éclatement existants. (Chapitres 2, 3 et 4)
- Unification et généralisation des méthodes inertielles et des moyennes de Mann. (Chapitres 2 et 3)
- Introduction de la notion de matrice superconcentrante qui permet l'obtention des résultats de sommabilité sur les algorithmes avec mémoire. Des exemples de telles matrices sont fournis. (Chapitre 3)
- Utilisation de la notion de matrice superconcentrante pour développer de nouveaux algorithmes avec mémoire. (Chapitre 3)

- Conception et étude des premiers algorithmes de point fixe multicouches sur-relaxés à métrique variable. (Chapitre 4)
- Conception et étude d'un algorithme explicite-implicite avec un pas proximal $(\gamma_n)_{n\in\mathbb{N}}$ pouvant dépasser la borne habituelle en contrepartie d'une sous-relaxation. (Chapitre 4)
- Calcul de nouveaux opérateurs proximaux. (Chapitre 5)
- Proposition d'utilisation d'opérateurs proximaux dans le cas de fonctions lisses. (Chapitre 5)
- Proposition d'un nouveau cadre pour les problèmes d'admissibilité convexe qui peut modéliser des contraintes lisses et non lisses. (Chapitre 5)
- Présentation de simulations numériques comparant des algorithmes implicites et d'autres explicites et montrant l'efficacité d'une activation implicite plutôt qu'explicite. (Chapitre 5)

1.4 Bibliographie

[1] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping,* Set-Valued Anal., 9 (2001), pp. 3–11.

[2] H. Attouch, A. Cabot, P. Frankel, and J. Peypouquet, *Alternating proximal algorithms for linearly constrained variational inequalities : Application to domain decomposition for PDE's,* Nonlinear Anal., 74 (2011), pp. 7455–7473.

[3] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than* $1/k^2$*,* SIAM J. Optim., 26 (2016), pp. 1824–1834.

[4] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, *Optimization with sparsity-inducing penalties,* Found. Trends Machine Learn., 4 (2012), pp. 1–106.

[5] J.-B. Baillon, R. E. Bruck, and S. Reich, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces,* Houston J. Math., 4 (1978), pp. 1–9.

[6] D. Barro and E. Canestrelli, *Combining stochastic programming and optimal control to decompose multistage stochastic optimization problems,* OR Spectrum, 38 (2016), pp. 711–742.

[7] H. H. Bauschke and J. M. Borwein, *On projection algorithms for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.

[8] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. Springer, New York, 2017.

[9] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems,* SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[10] R. I. Boţ, E. R. Csetnek, and C. Hendrich, *Inertial Douglas-Rachford splitting for monotone inclusion problems,* Appl. Math. Comput., 256 (2015), pp. 472–487.

[11] H. Brézis and P.-L. Lions, *Produits infinis de résolvantes,* Israel J. Math., 29 (1978), pp. 329–345.

[12] A. Cegielski, *Iterative Methods for Fixed Point Problems in Hilbert Spaces,* Lecture Notes in Mathematics, vol. 2057. Springer, Heidelberg, 2012.

[13] A. Chambolle and C. Dossal, *On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm",* J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[14] P. L. Combettes, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), pp. 475–504.

[15] P. L. Combettes and C. L. Müller, *Perspective functions : Proximal calculus and applications in high-dimensional statistics,* J. Math. Anal. Appl., 457 (2018), pp. 1283–1306.

[16] P. L. Combettes, *Monotone operator theory in convex optimization,* Math. Program. B170 (2018), pp. 177–206.

[17] P. L. Combettes and T. Pennanen, *Generalized Mann iterates for constructing fixed points in Hilbert spaces,* J. Math. Anal. Appl., 275 (2002), pp. 521–536.

[18] P. L. Combettes and J.-C. Pesquet, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,* IEEE J. Selected Topics Signal Process., 1 (2007), pp. 564–574.

[19] P. L. Combettes and J.-C. Pesquet, *Proximal splitting methods in signal processing,* in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* (H. H. Bauschke et al., eds.), pp. 185–212. Springer, New York, 2011.

[20] P. L. Combettes and J.-C. Pesquet, *Deep neural network structures solving variational inequalities,* arXiv preprint arXiv:1808.07526, 2018.

[21] P. L. Combettes, S. Salzo, and S. Villa, *Consistent learning by composite proximal thresholding,* Math. Program., B167 (2018), pp. 99–127.

[22] P. L. Combettes and B. C. Vũ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality,* Optimization, 63 (2014), pp. 1289–1318.

[23] P. L. Combettes and V. R. Wajs, *Signal recovery by proximal forward-backward splitting,* Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[24] P. L. Combettes and I. Yamada, *Compositions and convex combinations of averaged nonexpansive operators,* J. Math. Anal. Appl., 425 (2015), pp. 55–70.

[25] W. G. Dotson, *On the Mann iterative process,* Trans. Amer. Math. Soc., 149 (1970), pp. 65–73.

[26] J. Douglas and H. H. Rachford, *On the numerical solution of heat conduction problems in two or three space variables,* Trans. Amer. Math. Soc., 82 (1956), pp. 421–439.

[27] R. GLOWINSKI AND P. LE TALLEC, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics.* SIAM, Philadelphia, 1989.

[28] R. GLOWINSKI, S. J. OSHER, AND W. YIN (EDS.), *Splitting Methods in Communication, Imaging, Science, and Engineering.* Springer, New York, 2016.

[29] K. GOEBEL AND W. A. KIRK, *Topics in Metric Fixed Point Theory,* vol. 28. Cambridge University Press, Cambridge, 1990.

[30] D. GOLDFARB AND S. MA, *Fast multiple-splitting algorithms for convex optimization,* SIAM J. Optim., 22 (2012), pp. 533–556.

[31] P.-L. LIONS AND B. MERCIER, *Splitting algorithms for the sum of two nonlinear operators,* SIAM J. Numer. Anal., 16 (1979), 964–979.

[32] W. R. MANN, *Mean value methods in iteration,* Proc. Amer. Math. Soc., 4 (1953), pp. 506–510.

[33] W. R. MANN, *Averaging to improve convergence of iterative processes,* Lecture Notes in Math., 701 (1979), pp. 169–179.

[34] B. MERCIER, *Lectures on Topics in Finite Element Solution of Elliptic Problems* (Lectures on Mathematics, no. 63). Tata Institute of Fundamental Research, Bombay, 1979.

[35] A. MOUDAFI AND M. OLINY, *Convergence of a splitting inertial proximal method for monotone operators,* J. Comput. Appl. Math., 155 (2003), pp. 447–454.

[36] A. S. NEMIROVSKI AND D. B. YUDIN, *Problem Complexity and Method Efficiency in Optimization.* Wiley, New York, 1983.

[37] Y. NESTEROV, *A method for solving the convex programming problem with convergence rate $O(1/k^2)$,* Dokl. Akad. Nauk, 269 (1983), pp. 543–547.

[38] C. OUTLAW AND C. W. GROETSCH, *Averaging iteration in a Banach space,* Bull. Amer. Math. Soc., 75 (1969), pp. 430–432.

[39] N. PAPADAKIS, G. PEYRÉ, AND E. OUDET, *Optimal transport with proximal splitting,* SIAM J. Imaging Sci., 7 (2014), pp. 212–238.

[40] B. T. POLYAK, *Some methods of speeding up the convergence of iteration methods,* USSR Comput. Math. Math. Phys., 4 (1964), pp. 1–17.

[41] R. T. ROCKAFELLAR, *Monotone operators and the proximal point algorithm,* SIAM J. Control Optim., 14 (1976), pp. 877–898.

[42] S. SRA, S. NOWOZIN, AND S. J. WRIGHT, *Optimization for Machine Learning.* MIT Press, Cambridge, MA, 2012.

[43] P. TSENG, *Applications of a splitting algorithm to decomposition in convex programming and variational inequalities,* SIAM J. Control Optim., 29 (1991), pp. 119–138.

[44] B. C. VŨ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math., 38 (2013), pp. 667–681.

[45] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications I – Fixed-Point Theorems.* Springer-Verlag, New York, 1990.

[46] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications II/A – Linear Monotone Operators.* Springer-Verlag, New York, 1990.

[47] E. ZEIDLER, *Nonlinear Functional Analysis and Its Applications II/B – Nonlinear Monotone Operators.* Springer-Verlag, New York, 1990.

Chapitre 2

Itérations quasicontractantes : des moyennes de Mann aux méthodes inertielles

Dans ce chapitre, nous présentons un nouveau modèle itératif où la mise à jour est obtenue en appliquant une composition de quasicontractions à un point de l'enveloppe affine de l'orbite générée par les itérés précédents. Cette démarche permet d'unifier de nombreuses constructions algorithmiques incluant les moyennes de Mann, les méthodes inertielles et les méthodes multicouches. Elle fournit un cadre qui permet de proposer de nouveaux algorithmes pour résoudre des problèmes d'inclusion monotone et de minimisation convexe.

2.1 Description et résultats principaux

Dans ce chapitre nous nous concentrons sur le problème qui suit. C'est une formulation qui permet de couvrir un grand nombre d'applications classiques et dans des domaines très divers.

Problème 2.1 Soit m un entier strictement positif. Pour tout $n \in \mathbb{N}$ et tout $i \in \{1, \ldots, m\}$, soit $\alpha_{i,n} \in \,]0,1]$ et soit $T_{i,n} \colon \mathcal{H} \to \mathcal{H}$ un opérateur $\alpha_{i,n}$-moyenné si $i < m$, et $\alpha_{m,n}$-quasimoyenné si $i = m$. De plus, pour tout $n \in \mathbb{N}$, l'une des propriétés suivantes est satisfaite :

(a) $T_{m,n}$ est $\alpha_{m,n}$-moyenné.

(b) $m > 1$, $\alpha_{m,n} < 1$, et $\bigcap_{i=1}^{m} \operatorname{Fix} T_{i,n} \neq \varnothing$.

(c) $m = 1$.

L'objectif est de

$$\text{trouver}\quad x \in S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad \text{où} \quad (\forall n \in \mathbb{N}) \quad T_n = T_{1,n}\cdots T_{m,n}. \tag{2.1}$$

Algorithme 2.2 On considère le cadre du Problème 2.1. Pour tout $n \in \mathbb{N}$, soit ϕ_n une constante de moyennage de T_n, soit $\lambda_n \in \left]0, 1/\phi_n\right]$ et pour tout $i \in \{1, \ldots, m\}$, soit $e_{i,n} \in \mathcal{H}$. Soit $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ une matrice de réels qui satisfait les propriétés suivantes :

(a) $\sup_{n\in\mathbb{N}} \sum_{j=0}^{n} |\mu_{n,j}| < +\infty$.

(b) $(\forall n \in \mathbb{N})\ \sum_{j=0}^{n} \mu_{n,j} = 1$.

(c) $(\forall j \in \mathbb{N})\ \lim\limits_{n\to+\infty} \mu_{n,j} = 0$.

(d) Il existe une suite $(\chi_n)_{n\in\mathbb{N}}$ à valeurs dans $\left]0, +\infty\right[$ telle que $\inf_{n\in\mathbb{N}} \chi_n > 0$ et telle que toute suite $(\xi_n)_{n\in\mathbb{N}}$ à valeurs dans $\left[0, +\infty\right[$ qui satisfait

$$\Big(\exists\, (\varepsilon_n)_{n\in\mathbb{N}} \in \left[0, +\infty\right[^{\mathbb{N}}\Big) \quad \begin{cases} \sum_{n\in\mathbb{N}} \chi_n \varepsilon_n < +\infty \\ (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n \end{cases} \tag{2.2}$$

converge.

Soit $x_0 \in \mathcal{H}$. On itère

$$\begin{array}{l} \text{pour } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\ x_{n+1} = \overline{x}_n + \lambda_n\Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n + e_{m,n}) + e_{m-1,n}\cdots\big) + e_{2,n}\Big) + e_{1,n} - \overline{x}_n\Big). \end{array} \right. \end{array} \tag{2.3}$$

Le théorème principal de ce chapitre est le suivant.

Théorème 2.3 *On considère le cadre de l'Algorithme 2.2. Pour tout $n \in \mathbb{N}$, on définit*

$$\vartheta_n = \lambda_n \sum_{i=1}^{m} \|e_{i,n}\| \quad \textit{et} \quad (\forall i \in \{1, \ldots, m\}) \quad T_{i+,n} = \begin{cases} T_{i+1,n}\cdots T_{m,n}, & \textit{si } i \neq m; \\ \mathrm{Id}\,, & \textit{si } i = m, \end{cases} \tag{2.4}$$

et on pose

$$\nu_n \colon S \to \left[0, +\infty\right[: x \mapsto \vartheta_n\big(2\|\overline{x}_n - x\| + \vartheta_n\big). \tag{2.5}$$

Alors les propriétés suivantes sont satisfaites :

(i) *Soient $n \in \mathbb{N}$ et $x \in S$. Alors $\|x_{n+1} - x\| \leqslant \sum_{j=0}^{n} |\mu_{n,j}|\, \|x_j - x\| + \vartheta_n$.*

(ii) *Soient $n \in \mathbb{N}$ et $x \in S$. Alors*

$$\begin{aligned} \|x_{n+1} - x\|^2 \leqslant \sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 - \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n} \mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2 \\ - \lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 + \nu_n(x). \end{aligned}$$

(iii) *Soient* $n \in \mathbb{N}$ *et* $x \in S$. *Alors*

$$\begin{aligned}\|x_{n+1} - x\|^2 \leqslant & \sum_{j=0}^{n} \mu_{n,j} \|x_j - x\|^2 - \frac{1}{2} \sum_{j=0}^{n} \sum_{k=0}^{n} \mu_{n,j}\mu_{n,k} \|x_j - x_k\|^2 \\ & + \lambda_n(\lambda_n - 1)\|T_n\overline{x}_n - \overline{x}_n\|^2 \\ & - \lambda_n \max_{1\leqslant i\leqslant m} \left(\frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2 \right) \\ & + \nu_n(x).\end{aligned}$$

Supposons que

$$\sum_{n\in\mathbb{N}} \chi_n \sum_{j=0}^{n} \sum_{k=0}^{n} [\mu_{n,j}\mu_{n,k}]^- \|x_j - x_k\|^2 < +\infty \quad \textit{et} \quad (\forall x \in S) \quad \sum_{n\in\mathbb{N}} \chi_n \nu_n(x) < +\infty. \tag{2.6}$$

Alors les propriétés suivantes sont également satisfaites :

(iv) *Soit* $x \in S$. *Alors* $(\|x_n - x\|)_{n\in\mathbb{N}}$ *converge.*

(v) $\lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 \to 0$.

(vi) $\sum_{j=0}^{n} \sum_{k=0}^{n} [\mu_{n,j}\mu_{n,k}]^+ \|x_j - x_k\|^2 \to 0$.

(vii) *Supposons que*

$$(\exists\, \varepsilon \in \,]0,1[)(\forall n \in \mathbb{N}) \quad \lambda_n \leqslant (1-\varepsilon)/\phi_n. \tag{2.7}$$

Alors $x_{n+1} - \overline{x}_n \to 0$. *De plus, si tout point d'accumulation séquentiel faible de* $(\overline{x}_n)_{n\in\mathbb{N}}$ *est dans* S, *alors il existe* $x \in S$ *tel que* $x_n \rightharpoonup x$.

(viii) *Supposons que* $(\overline{x}_n)_{n\in\mathbb{N}}$ *a un point d'accumulation fort* x *dans* S *et que* (2.7) *est satisfaite. Alors* $x_n \to x$.

(ix) *Soit* $x \in S$. *Supposons que* $(\exists\, \varepsilon \in \,]0,1[)(\forall n \in \mathbb{N})$ $\lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n$. *Alors*

$$\lambda_n \max_{1\leqslant i\leqslant m} \frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2 \to 0.$$

Ci-dessous on présente un corollaire du Théorème 2.3 pour une certaine classe de matrices. Il permet de développer des algorithmes dits inertiels.

Corollaire 2.4 *On considère le cadre de l'Algorithme 2.2 avec* $(\forall i \in \{1,\ldots,m\})(\forall n \in \mathbb{N})$ $e_{i,n} = 0$. *On pose* $x_{-1} = x_0$ *et on suppose qu'il existe une suite* $(\eta_n)_{n\in\mathbb{N}}$ *à valeurs dans* $[0,1[$ *telle que* $\eta_0 = 0$ *et*

$$(\forall n \in \mathbb{N})(\forall j \in \{0,\ldots,n\}) \quad \mu_{n,j} = \begin{cases} 1+\eta_n, & \textit{si } \ j = n; \\ -\eta_n, & \textit{si } \ j = n-1; \\ 0, & \textit{si } \ j < n-1. \end{cases} \tag{2.8}$$

Pour tout $n \in \mathbb{N}$, on pose

$$(\forall k \in \mathbb{N}) \quad \zeta_{k,n} = \begin{cases} 0, & \textit{si } \ k \leqslant n; \\ \displaystyle\sum_{j=n+1}^{k} (\eta_j - 1), & \textit{si } \ k > n, \end{cases} \tag{2.9}$$

et on suppose que $\chi_n = \sum_{k \geqslant n} \exp(\zeta_{k,n})$. *Supposons que l'une des propriétés suivantes est satisfaite :*

(a) $\sum_{n \in \mathbb{N}} \chi_n \eta_n \|x_n - x_{n-1}\|^2 < +\infty$.

(b) $\sum_{n \in \mathbb{N}} n \|x_n - x_{n-1}\|^2 < +\infty$ *et il existe* $\tau \in [2, +\infty[$ *tel que* $(\forall n \in \mathbb{N} \smallsetminus \{0\})$ $\eta_n = (n-1)/(n+\tau)$.

(c) $\sum_{n \in \mathbb{N}} \eta_n \|x_n - x_{n-1}\|^2 < +\infty$ *et il existe* $\eta \in [0, 1[$ *tel que* $(\forall n \in \mathbb{N})$ $\eta_n \leqslant \eta$.

(d) *On pose* $(\forall n \in \mathbb{N})$ $\omega_n = 1/\phi_n - \lambda_n$. *Il existe* $(\sigma, \vartheta) \in]0, +\infty[^2$ *et* $\eta \in]0, 1[$ *tels que*

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \eta_n \leqslant \eta_{n+1} \leqslant \eta \\ \lambda_n \leqslant \dfrac{\vartheta/\phi_n - \eta\big(\eta(1+\eta) + \eta\vartheta\omega_{n+1} + \sigma\big)}{\vartheta\big(1 + \eta(1+\eta) + \eta\vartheta\omega_{n+1} + \sigma\big)} \\ \dfrac{\eta^2(1+\eta) + \eta\sigma}{\vartheta} < \dfrac{1}{\phi_n} - \eta^2\omega_{n+1}. \end{cases} \tag{2.10}$$

Alors les propriétés suivantes sont satisfaites :

(i) $\lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 \to 0$.

(ii) *Soit* $x \in S$. *Supposons que* $(\exists\, \varepsilon \in]0, 1[)(\forall n \in \mathbb{N})$ $\lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n$. *Alors*

$$\lambda_n \max_{1 \leqslant i \leqslant m} \frac{1 - \alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2 \to 0.$$

(iii) $\overline{x}_n - x_n \to 0$.

(iv) *Supposons que tout point d'accumulation séquentiel faible de* $(\overline{x}_n)_{n \in \mathbb{N}}$ *est dans* S. *Alors il existe* $x \in S$ *tel que* $x_n \rightharpoonup x$.

Remarque 2.5 Le cas (d) avec $m = 1$ et une contraction correspond aux résultats de [14]. Ce résultat couvre aussi le cas d'un opérateur moyenné puisque l'itération

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n(Tx_n - x_n) \tag{2.11}$$

avec $\alpha \in]0, 1]$ et $T\colon \mathcal{H} \to \mathcal{H}$ un opérateur α-moyenné peut être ramené à

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\alpha(Rx_n - x_n) \tag{2.12}$$

où $R\colon \mathcal{H} \to \mathcal{H}$ est une contraction. Ainsi, afin de traiter le cas des opérateurs α-moyennés, il est possible d'utiliser [14]. Néanmoins, cela donne un résultat moins précis que lorsque l'on passe directement par les opérateurs moyennés : la borne supérieure de la suite $(\lambda_n)_{n \in \mathbb{N}}$ est plus petite.

Nous proposons une nouvelle méthode explicite-implicite avec mémoire.

Proposition 2.6 *Soit $\beta \in \left]0,+\infty\right[$, soit $\varepsilon \in \left]0,\min\{1/2,\beta\}\right[$, soit $x_0 \in \mathcal{H}$, soit $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ un opérateur maximalement monotone et soit $B\colon \mathcal{H} \to \mathcal{H}$ un opérateur β-cocoercif. Soit $(\gamma_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $[\varepsilon, 2\beta/(1+\varepsilon)]$, soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels qui satisfait les conditions (a)–(d) de l'Algorithme 2.2, et soient $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites à valeurs dans $\mathcal{H}$ telles que $\sum_{n\in\mathbb{N}} \chi_n\|a_n\| < +\infty$ et $\sum_{n\in\mathbb{N}} \chi_n\|b_n\| < +\infty$. Supposons que* zer $(A+B) \neq \varnothing$ *et que*

$$(\forall n \in \mathbb{N}) \quad \lambda_n \in \left[\varepsilon, 1+(1-\varepsilon)\left(1-\frac{\gamma_n}{2\beta}\right)\right], \tag{2.13}$$

et on pose $(\forall n \in \mathbb{N})$ $\phi_n = 2/(4-\gamma_n/\beta)$. Pour tout $n \in \mathbb{N}$, on itère

$$x_{n+1} = \overline{x}_n + \lambda_n\Big(J_{\gamma_n A}\big(\overline{x}_n - \gamma_n(B\overline{x}_n + b_n)\big) + a_n - \overline{x}_n\Big), \quad \text{où} \quad \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j}x_j. \tag{2.14}$$

Supposons que l'une des propriétés suivantes est satisfaite :

(a) $\inf\limits_{n\in\mathbb{N}} \min\limits_{0\leqslant j\leqslant n} \mu_{n,j} \geqslant 0$.

(b) $a_n \equiv b_n \equiv 0$, $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *satisfait* (2.8), *et l'une des conditions (a)–(d) dans le Corollaire 2.4 est satisfaite.*

Alors les propriétés suivantes sont satisfaites :

(i) $J_{\gamma_n A}(\overline{x}_n - \gamma_n B\overline{x}_n) - \overline{x}_n \to 0$.

(ii) *Soit* $z \in$ zer $(A+B)$. *Alors* $B\overline{x}_n \to Bz$.

(iii) *Il existe* $x \in$ zer $(A+B)$ *tel que* $x_n \rightharpoonup x$.

(iv) *Supposons que A ou B est demi-régulier en tout point de* zer $(A+B)$. *Alors il existe* $x \in$ zer $(A+B)$ *tel que* $x_n \to x$.

2.2 Article en anglais

QUASINONEXPANSIVE ITERATIONS ON THE AFFINE HULL OF ORBITS : FROM MANN'S MEAN VALUE ALGORITHM TO INERTIAL METHODS [1]

Abstract. Fixed point iterations play a central role in the design and the analysis of a large number of optimization algorithms. We study a new iterative scheme in which

1. P. L. Combettes and L. E. Glaudin, *Quasinonexpansive iterations on the affine hull of orbits : From Mann's mean value algorithm to inertial methods,* SIAM J. Optim., 27 (2017), pp. 2356–2380.

the update is obtained by applying a composition of quasinonexpansive operators to a point in the affine hull of the orbit generated up to the current iterate. This investigation unifies several algorithmic constructs, including Mann's mean value method, inertial methods, and multi-layer memoryless methods. It also provides a framework for the development of new algorithms, such as those we propose for solving monotone inclusion and minimization problems.

Keywords. Averaged operator, fixed point iteration, forward-backward algorithm, inertial algorithm, mean value iterations, monotone operator splitting, nonsmooth minimization, Peaceman-Rachford algorithm, proximal algorithm.

AMS 2010 Subject Classification : Primary 65J15, 47H09 ; Secondary 47H05, 65K05, 90C25

Dedicated to the memory of Felipe Álvarez 1972-2017

2.2.1 Introduction

Algorithms arising in various branches of optimization can be efficiently modeled and analyzed as fixed point iterations in a real Hilbert space $\mathcal{H}$; see, e.g., [9, 10, 13, 16, 18, 19, 22, 26, 43]. Our paper unifies three important algorithmic fixed point frameworks that coexist in the literature : mean value methods, inertial methods, and multilayer memoryless methods.

Let $T\colon \mathcal{H} \to \mathcal{H}$ be an operator with fixed point set $\operatorname{Fix} T$. In 1953, inspired by classical results on the summation of divergent series [11, 29, 44], Mann [34] proposed to extend the standard successive approximation scheme

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = Tx_n \tag{2.15}$$

to the mean value algorithm

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = T\overline{x}_n, \quad \text{where} \quad \overline{x}_n \in \operatorname{conv}\,(x_j)_{0\leqslant j\leqslant n}. \tag{2.16}$$

In other words, the operator T is not applied to the most current iterate as in the memoryless (single step) process (2.15), but to a point in the convex hull of the orbit $(x_j)_{0\leqslant j\leqslant n}$ generated so far. His motivation was that, although the sequence generated by (2.15) may fail to converge to a fixed point of T, that generated by (2.16) can under suitable conditions. This work was followed by interesting developments and analyses of such mean value iterations, e.g., [8, 12, 15, 28, 30, 32, 35, 37, 42], especially in the case when T is nonexpansive (1-Lipschitzian) or merely quasinonexpansive, that is (this notion was essentially introduced in [27])

$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{Fix} T) \quad \|Tx - y\| \leqslant \|x - y\|. \tag{2.17}$$

In [21], the asymptotic behavior of the mean value process

$$x_0 \in \mathcal{H} \text{ and } (\forall n \in \mathbb{N}) \quad x_{n+1} = \overline{x}_n + \lambda_n\big(T_n\overline{x}_n + e_n - \overline{x}_n\big), \text{ where } \overline{x}_n \in \operatorname{conv}\big(x_j\big)_{0\leqslant j\leqslant n}, \tag{2.18}$$

was investigated under general conditions on the construction of the averaging process $(\overline{x}_n)_{n\in\mathbb{N}}$ and the assumptions that, for every $n \in \mathbb{N}$, $e_n \in \mathcal{H}$ models a possible error made in the computation of $T_n\overline{x}_n$, $\lambda_n \in \,]0,2[$, and $T_n\colon \mathcal{H} \to \mathcal{H}$ is firmly quasinonexpansive, i.e., $2T_n - \mathrm{Id}$ is quasinonexpansive or, equivalently [10],

$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{Fix} T_n) \quad \langle y - T_n x \mid x - T_n x\rangle \leqslant 0. \tag{2.19}$$

The idea of using the past of the orbit generated by an algorithm can also be found in the work of Polyak [39, 41], who drew inspiration from classical multistep methods in numerical analysis. His motivation was to improve the speed of convergence over memoryless methods. For instance, the classical gradient method [38] for minimizing a smooth convex function $f\colon \mathcal{H} \to \mathbb{R}$ is an explicit discretization of the continuous-time process $-\dot{x}(t) = \nabla f(x(t))$. Polyak [39] proposed to consider instead the process $-\ddot{x}(t) - \beta\dot{x}(t) = \nabla f(x(t))$, where $\beta \in \,]0,+\infty[$, and studied the algorithm resulting from its explicit discretization. He observed that, from a mechanical viewpoint, the term $\ddot{x}(t)$ can be interpreted as an inertial component. More generally, for a proper lower semicontinuous convex function $f\colon \mathcal{H} \to \,]-\infty,+\infty]$, Álvarez investigated in [1] an implicit discretization of the inertial differential inclusion $-\ddot{x}(t) - \beta\dot{x}(t) \in \partial f(x(t))$, namely

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma_n f}\overline{x}_n + e_n, \quad \text{where} \quad \begin{cases} \overline{x}_n = (1+\eta_n)x_n - \eta_n x_{n-1} \\ \eta_n \in [0,1[\\ \gamma_n \in \,]0,+\infty[\,, \end{cases} \tag{2.20}$$

and where prox_f is the proximity operator of f [10, 36]. The inertial proximal point algorithm (2.20) has been extended in various directions, e.g., [3, 14, 17] ; see also [5] for further motivation in the context of nonconvex minimization problems.

Working from a different perspective, a structured extension of (2.15) involving the composition of m averaged nonexpansive operators was proposed in [19]. This m-layer algorithm is governed by the memoryless recursion

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\big(T_{1,n}\cdots T_{m,n}x_n + e_n - x_n\big), \quad \text{where} \quad \lambda_n \in \,]0,1]\,. \tag{2.21}$$

Recall that a nonexpansive operator $T\colon \mathcal{H} \to \mathcal{H}$ is averaged with constant $\alpha \in \,]0,1[$ if there exists a nonexpansive operator $R\colon \mathcal{H} \to \mathcal{H}$ such that $T = (1-\alpha)\,\mathrm{Id} + \alpha R$ [7, 10]. The multilayer iteration process (2.21) was shown in [19] to provide a synthetic analysis of various algorithms, in particular in the area of monotone operator splitting methods. It was extended in [25] to an overrelaxed method, i.e., one with parameters $(\lambda_n)_{n\in\mathbb{N}}$ possibly larger than 1.

In the literature, the asymptotic analysis of the above methods has been carried out independently because of their apparent lack of common structure. In the present paper, we exhibit a structure that unifies (2.15), (2.16), (2.18), (2.20), and (2.21) in a single algorithm of the form

$$x_0 \in \mathcal{H} \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = \overline{x}_n + \lambda_n\big(T_{1,n}\cdots T_{m,n}\overline{x}_n + e_n - \overline{x}_n\big),$$
$$\text{where} \quad \overline{x}_n \in \operatorname{aff}\,(x_j)_{0\leqslant j\leqslant n} \quad \text{and} \quad \lambda_n \in \left]0,+\infty\right[, \tag{2.22}$$

under the assumption that each operator $T_{i,n}$ is $\alpha_{i,n}$-averaged quasinonexpansive, i.e.,

$$(\forall x \in \mathcal{H})(\forall y \in \operatorname{Fix} T_{i,n})$$
$$2(1-\alpha_{i,n})\langle y - T_{i,n}x \mid x - T_{i,n}x\rangle \leqslant (2\alpha_{i,n}-1)\big(\|x-y\|^2 - \|T_{i,n}x - y\|^2\big), \tag{2.23}$$

for some $\alpha_{i,n} \in \left]0,1\right]$, which means that the operator $(1-1/\alpha_{i,n})\operatorname{Id} + (1/\alpha_{i,n})T_{i,n}$ is quasinonexpansive. In words, at iteration n, a point $\overline{x}_n$ is picked in the affine hull of the orbit $(x_j)_{0\leqslant j\leqslant n}$ generated so far, a composition of quasinonexpansive operators is applied to it, up to some error e_n, and the update x_{n+1} is obtained via a relaxation with parameter λ_n. Note that (2.22)–(2.23) not only brings together mean value iterations, inertial methods, and the memoryless multilayer setting of [19, 25], but also provides a flexible framework to design new iterative methods.

The fixed point problem under consideration will be the following (note that we allow 1 as an averaging constant for added flexibility).

Problem 2.7 Let m be a strictly positive integer. For every $n \in \mathbb{N}$ and every $i \in \{1,\ldots,m\}$, $\alpha_{i,n} \in \left]0,1\right]$ and $T_{i,n}\colon \mathcal{H} \to \mathcal{H}$ is $\alpha_{i,n}$-averaged nonexpansive if $i < m$, and $\alpha_{m,n}$-averaged quasinonexpansive if $i = m$. In addition,

$$S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad \text{where} \quad (\forall n \in \mathbb{N}) \quad T_n = T_{1,n}\cdots T_{m,n}, \tag{2.24}$$

and one of the following holds :

(a) For every $n \in \mathbb{N}$, $T_{m,n}$ is $\alpha_{m,n}$-averaged nonexpansive.
(b) $m > 1$ and, for every $n \in \mathbb{N}$, $\alpha_{m,n} < 1$ and $\bigcap_{i=1}^m \operatorname{Fix} T_{i,n} \neq \varnothing$.
(c) $m = 1$.

The problem is to find a point in S.

To solve Problem 2.7, we are going to employ (2.22), which we now formulate more formally.

Algorithm 2.8 Consider the setting of Problem 2.7. For every $n \in \mathbb{N}$, let ϕ_n be an averaging constant of T_n, let $\lambda_n \in \left]0, 1/\phi_n\right]$ and, for every $i \in \{1,\ldots,m\}$, let $e_{i,n} \in \mathcal{H}$. Let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be a real array which satisfies the following :

(a) $\sup_{n\in\mathbb{N}} \sum_{j=0}^{n} |\mu_{n,j}| < +\infty$.

(b) $(\forall n \in \mathbb{N})\ \sum_{j=0}^{n} \mu_{n,j} = 1$.

(c) $(\forall j \in \mathbb{N})\ \lim_{n\to+\infty} \mu_{n,j} = 0$.

(d) There exists a sequence $(\chi_n)_{n\in\mathbb{N}}$ in $]0,+\infty[$ such that $\inf_{n\in\mathbb{N}} \chi_n > 0$ and every sequence $(\xi_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ that satisfies

$$\Big(\exists\, (\varepsilon_n)_{n\in\mathbb{N}} \in [0,+\infty[^{\mathbb{N}}\Big) \quad \begin{cases} \sum_{n\in\mathbb{N}} \chi_n \varepsilon_n < +\infty \\ (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n \end{cases} \tag{2.25}$$

converges.

Let $x_0 \in \mathcal{H}$ and set

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\ x_{n+1} = \overline{x}_n + \lambda_n \Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n + e_{m,n}) + e_{m-1,n} \cdots\big) + e_{2,n}\Big) + e_{1,n} - \overline{x}_n\Big). \end{array} \right. \end{array} \tag{2.26}$$

Remark 2.9 Here are some comments about the parameters appearing in Problem 2.7 and Algorithm 2.8.

(i) The composite operator T_n of (2.24) is averaged quasinonexpansive with constant

$$\phi_n = \begin{cases} \left(1 + \left(\displaystyle\sum_{i=1}^{m} \frac{\alpha_{i,n}}{1-\alpha_{i,n}}\right)^{-1}\right)^{-1}, & \text{if } \displaystyle\max_{1\leqslant i\leqslant m} \alpha_{i,n} < 1; \\ 1, & \text{otherwise.} \end{cases} \tag{2.27}$$

The proof is given in [25, Proposition 2.5] for case (a) of Problem 2.7. It easily extends to case (b), while case (c) is trivial.

(ii) Examples of arrays $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ that satisfy conditions (a)–(d) in Algorithm 2.8 are provided in [21, Section 2] in the case of mean value iterations, i.e., $\inf_{n\in\mathbb{N}} \min_{0\leqslant j\leqslant n} \mu_{n,j} \geqslant 0$, with $\chi_n \equiv 1$. An important instance with negative coefficients will be presented in Example 2.14.

(iii) The term $e_{i,n}$ in (2.26) models a possible numerical error in the implementation of the operator $T_{i,n}$.

The material is organized as follows. In Section 2.2.2 we provide preliminary results. The main results on the convergence of the orbits of Algorithm 2.8 are presented in Section 2.2.3. Section 2.2.4 is dedicated to new algorithms for fixed point computation, monotone operator splitting, and nonsmooth minimization based on the proposed framework.

Notation. $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$ and associated norm $\|\cdot\|$. We denote by Id the identity operator on $\mathcal{H}$; $\rightharpoonup$ and $\to$ denote, respectively, weak and strong convergence in $\mathcal{H}$. The positive and negative parts of $\xi \in \mathbb{R}$ are, respectively, $\xi^+ = \max\{0, \xi\}$ and $\xi^- = -\min\{0, \xi\}$. Finally, $\delta_{n,j}$ is the Kronecker delta : it takes on the value 1 if $n = j$, and 0 otherwise.

2.2.2 Preliminary results

In this section we establish some technical facts that will be used subsequently. We start with a Grönwall-type result.

Lemma 2.10 *Let $(\theta_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be sequences in $[0, +\infty[$, and let $(\nu_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$ such that $(\forall n \in \mathbb{N})\ \theta_{n+1} \leqslant (1 + \nu_n)\theta_n + \varepsilon_n$. Then*

$$(\forall n \in \mathbb{N}) \quad \theta_{n+1} \leqslant \theta_0 \exp\left(\sum_{k=0}^{n} \nu_k\right) + \sum_{j=0}^{n-1} \varepsilon_j \exp\left(\sum_{k=j+1}^{n} \nu_k\right) + \varepsilon_n. \tag{2.28}$$

Proof. We have $(\forall n \in \mathbb{N})\ 1 + \nu_n \leqslant \exp(\nu_n)$. Therefore $\theta_1 \leqslant \theta_0 \exp(\nu_0) + \varepsilon_0$ and

$$\begin{aligned}(\forall n \in \mathbb{N} \smallsetminus \{0\}) \quad \theta_{n+1} &\leqslant \theta_n \exp(\nu_n) + \varepsilon_n \\ &\leqslant \theta_{n-1} \exp(\nu_n)\exp(\nu_{n-1}) + \varepsilon_{n-1}\exp(\nu_n) + \varepsilon_n \\ &\leqslant \theta_0 \prod_{k=0}^{n} \exp(\nu_k) + \sum_{j=0}^{n-1} \varepsilon_j \prod_{k=j+1}^{n} \exp(\nu_k) + \varepsilon_n \\ &= \theta_0 \exp\left(\sum_{k=0}^{n} \nu_k\right) + \sum_{j=0}^{n-1} \varepsilon_j \exp\left(\sum_{k=j+1}^{n} \nu_k\right) + \varepsilon_n, \end{aligned} \tag{2.29}$$

as claimed. □

Lemma 2.11 [31, Theorem 43.5] *Let $(\xi_n)_{n\in\mathbb{N}}$ be a sequence in $\mathbb{R}$, let $\xi \in \mathbb{R}$, suppose that $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ is a real array that satisfies conditions (a)–(c) in Algorithm 2.8. Then $\xi_n \to \xi \Rightarrow \sum_{j=0}^{n} \mu_{n,j}\xi_j \to \xi$.*

Lemma 2.12 *Let $(\beta_n)_{n\in\mathbb{N}}$, $(\gamma_n)_{n\in\mathbb{N}}$, $(\delta_n)_{n\in\mathbb{N}}$, $(\eta_n)_{n\in\mathbb{N}}$, and $(\lambda_n)_{n\in\mathbb{N}}$ be sequences in $[0, +\infty[$, let $(\phi_n)_{n\in\mathbb{N}}$ be a sequence in $]0, 1]$, let $(\vartheta, \sigma) \in\ ]0, +\infty[^2$, and let $\eta \in\ ]0, 1[$. Set $\beta_{-1} = \beta_0$ and*

$$(\forall n \in \mathbb{N}) \quad \omega_n = \frac{1}{\phi_n} - \lambda_n, \tag{2.30}$$

and suppose that the following hold :

(a) $(\forall n \in \mathbb{N}) \quad \eta_n \leqslant \eta_{n+1} \leqslant \eta.$

(b) $(\forall n \in \mathbb{N}) \quad \gamma_n \leqslant \eta(1+\eta) + \eta\vartheta\omega_n$.

(c) $(\forall n \in \mathbb{N}) \quad \dfrac{\eta^2(1+\eta)+\eta\sigma}{\vartheta} < \dfrac{1}{\phi_n} - \eta^2\omega_{n+1}$.

(d) $(\forall n \in \mathbb{N}) \quad 0 < \lambda_n \leqslant \dfrac{\vartheta/\phi_n - \eta\big(\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma\big)}{\vartheta\big(1+\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma\big)}$.

(e) $(\forall n \in \mathbb{N}) \quad \beta_{n+1}-\beta_n-\eta_n(\beta_n-\beta_{n-1}) \leqslant \dfrac{(1/\phi_n-\lambda_n)\big(\eta_n/(\eta_n+\vartheta\lambda_n)-1\big)}{\lambda_n}\delta_{n+1}+\gamma_n\delta_n$.

Then $\sum_{n\in\mathbb{N}}\delta_n < +\infty$.

Proof. We use arguments similar to those used in [3, 14]. It follows from (c) that $(\forall n \in \mathbb{N})$ $0 < \vartheta/\phi_n - \eta^2\omega_{n+1}\vartheta - \eta^2(1+\eta) - \eta\sigma$. This shows that $(\lambda_n)_{n\in\mathbb{N}}$ is well defined. Now set $(\forall n \in \mathbb{N})$ $\rho_n = 1/(\eta_n+\vartheta\lambda_n)$ and $\kappa_n = \beta_n - \eta_n\beta_{n-1} + \gamma_n\delta_n$. We derive from (a) and (e) that

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad \kappa_{n+1}-\kappa_n &\leqslant \beta_{n+1}-\eta_n\beta_n-\beta_n+\eta_n\beta_{n-1}+\gamma_{n+1}\delta_{n+1}-\gamma_n\delta_n \\
&\leqslant \left(\frac{(1/\phi_n-\lambda_n)(\eta_n\rho_n-1)}{\lambda_n}+\gamma_{n+1}\right)\delta_{n+1}.
\end{aligned}
\tag{2.31}
$$

On the other hand, $(\forall n \in \mathbb{N})$ $\vartheta(1+(\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma)) > 0$. Consequently, (d) can be written as

$$
(\forall n \in \mathbb{N}) \quad \vartheta\lambda_n+\vartheta\lambda_n\big(\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma\big) \leqslant \frac{\vartheta}{\phi_n} - \eta\big(\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma\big). \tag{2.32}
$$

Using (a) and (b), and then (2.32), we get

$$
(\forall n \in \mathbb{N}) \quad (\eta_n+\vartheta\lambda_n)(\gamma_{n+1}+\sigma)+\vartheta\lambda_n \leqslant (\eta+\vartheta\lambda_n)\big(\eta(1+\eta)+\eta\vartheta\omega_{n+1}+\sigma\big)+\vartheta\lambda_n \leqslant \frac{\vartheta}{\phi_n}. \tag{2.33}
$$

However,

$$
\begin{aligned}
(\forall n \in \mathbb{N}) \quad & (\eta_n+\vartheta\lambda_n)(\gamma_{n+1}+\sigma)+\vartheta\lambda_n \leqslant \frac{\vartheta}{\phi_n} \\
&\Leftrightarrow (\eta_n+\vartheta\lambda_n)(\gamma_{n+1}+\sigma)-(1/\phi_n-\lambda_n)\vartheta \leqslant 0 \\
&\Leftrightarrow (1/\phi_n-\lambda_n)\left(\frac{-\vartheta}{\eta_n+\vartheta\lambda_n}\right) \leqslant -(\gamma_{n+1}+\sigma) \\
&\Leftrightarrow \frac{(1/\phi_n-\lambda_n)(\eta_n\rho_n-1)}{\lambda_n}+\gamma_{n+1} \leqslant -\sigma.
\end{aligned}
\tag{2.34}
$$

It therefore follows from (2.31) and (2.33) that

$$
(\forall n \in \mathbb{N}) \quad \kappa_{n+1}-\kappa_n \leqslant -\sigma\delta_{n+1}. \tag{2.35}
$$

Thus, $(\kappa_n)_{n\in\mathbb{N}}$ is decreasing and

$$(\forall n \in \mathbb{N}) \quad \beta_n - \eta\beta_{n-1} = \kappa_n - \gamma_n\delta_n \leqslant \kappa_n \leqslant \kappa_0, \tag{2.36}$$

from which we infer that $(\forall n \in \mathbb{N})\ \beta_n \leqslant \kappa_0 + \eta\beta_{n-1}$. In turn,

$$(\forall n \in \mathbb{N} \smallsetminus \{0\}) \quad \beta_n \leqslant \eta^n\beta_0 + \kappa_0 \sum_{j=0}^{n-1} \eta^j \leqslant \eta^n\beta_0 + \frac{\kappa_0}{1-\eta}. \tag{2.37}$$

Altogether, we derive from (2.35), (2.36), and (2.37) that

$$(\forall n \in \mathbb{N}) \quad \sigma \sum_{j=0}^{n} \delta_{j+1} \leqslant \kappa_0 - \kappa_{n+1} \leqslant \kappa_0 + \eta\beta_n \leqslant \frac{\kappa_0}{1-\eta} + \eta^{n+1}\beta_0. \tag{2.38}$$

Hence, $\sum_{j\geqslant 1} \delta_j \leqslant \kappa_0/((1-\eta)\sigma) < +\infty$, and the proof is complete. ☐

Lemma 2.13 *Let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1[$. For every $n \in \mathbb{N}$, set*

$$(\forall k \in \mathbb{N}) \quad \zeta_{k,n} = \begin{cases} 0, & \text{if } k \leqslant n; \\ \displaystyle\sum_{j=n+1}^{k} (\eta_j - 1), & \text{if } k > n, \end{cases} \tag{2.39}$$

and $\chi_n = \sum_{k\geqslant n} \exp(\zeta_{k,n})$. Then the following hold :

(i) *Let $\tau \in [2,+\infty[$ and suppose that $(\forall n \in \mathbb{N})\ \eta_{n+1} = n/(n+1+\tau)$. Then $(\forall n \in \mathbb{N})$ $\chi_n \leqslant (n+7)/2$.*

(ii) *Suppose that $(\exists\, \eta \in [0,1[)(\forall n \in \mathbb{N})\ \eta_n \leqslant \eta$. Then $(\forall n \in \mathbb{N})\ \chi_n \leqslant e/(1-\eta)$.*

Proof. (i) : We have $(\forall n \in \mathbb{N})(\forall k \in \{n+1, n+2, \ldots\})\ \zeta_{k,n} = -(1+\tau)\sum_{j=n+1}^{k} 1/(j+\tau) \leqslant -3\sum_{j=n+1}^{k} 1/(j+2)$. Since $\xi \mapsto 1/(\xi+2)$ is decreasing on $[1,+\infty[$, it follows that

$$(\forall n \in \mathbb{N})(\forall k \in \{n+1, n+2, \ldots\}) \quad \zeta_{k,n} \leqslant -3\int_{n+1}^{k+1} \frac{d\xi}{\xi+2} = \ln \frac{(n+3)^3}{(k+3)^3}. \tag{2.40}$$

Furthermore, since $\xi \mapsto 1/(\xi+3)^3$ is decreasing on $]-1,+\infty[$, (2.39) yields

$$(\forall n \in \mathbb{N}) \quad \chi_n \leqslant \sum_{k\geqslant n} \frac{(n+3)^3}{(k+3)^3} \leqslant (n+3)^3 \int_{n-1}^{+\infty} \frac{d\xi}{(\xi+3)^3} = \frac{(n+3)^3}{2(n+2)^2} \leqslant \frac{n+7}{2}. \tag{2.41}$$

(ii) : Note that

$$(\forall n \in \mathbb{N})(\forall k \in \{n+1, n+2, \ldots\}) \quad \zeta_{k,n} = \sum_{j=n+1}^{k} (\eta_j - 1) \leqslant \sum_{j=n+1}^{k} (\eta - 1) = (\eta-1)(k-n).$$

(2.42)

Since $\xi \mapsto \exp((\eta-1)\xi)$ is decreasing on $]-1,+\infty[$, it follows that

$$(\forall n \in \mathbb{N}) \quad \chi_n \leqslant \sum_{k \geqslant n} \exp\big((\eta-1)(k-n)\big) \leqslant \int_{n-1}^{+\infty} \exp\big((\eta-1)(\xi-n)\big)d\xi = \frac{\exp(1-\eta)}{1-\eta}, \tag{2.43}$$

which proves the assertion. □

The next example provides an instance of an array $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ satisfying the conditions of Algorithm 2.8 with negative entries. This example will be central to the study of the convergence of some inertial methods.

Example 2.14 Let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be a real array such that $\mu_{0,0} = 1$ and

$$(\forall n \in \mathbb{N}) \quad 1 \leqslant \mu_{n,n} < 2 \quad \text{and} \quad (\forall j \in \{0,\ldots,n\}) \quad \mu_{n,j} = \begin{cases} 1-\mu_{n,n}, & \text{if } j = n-1; \\ 0, & \text{if } j < n-1. \end{cases} \tag{2.44}$$

For every $n \in \mathbb{N}$, set

$$(\forall k \in \mathbb{N}) \quad \zeta_{k,n} = \begin{cases} 0, & \text{if } k \leqslant n; \\ \displaystyle\sum_{j=n+1}^{k} (\mu_{j,j}-2), & \text{if } k > n, \end{cases} \tag{2.45}$$

and suppose that $\chi_n = \sum_{k\geqslant n} \exp(\zeta_{k,n}) < +\infty$. Then $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ satisfies conditions (a)–(d) in Algorithm 2.8.

Proof. (a) : $(\forall n \in \mathbb{N})$ $\sum_{j=0}^{n} |\mu_{n,j}| = \mu_{n,n} + |1-\mu_{n,n}| \leqslant 3$.

(b) : $(\forall n \in \mathbb{N})$ $\sum_{j=0}^{n} \mu_{n,j} = (1-\mu_{n,n}) + \mu_{n,n} = 1$.

(c) : Let $j \in \mathbb{N}$. Then $(\forall n \in \mathbb{N})$ $n > j+1 \Rightarrow \mu_{n,j} = 0$. Hence, $\lim\limits_{n\to+\infty} \mu_{n,j} = 0$.

(d) : We have $(\forall n \in \mathbb{N})$ $\chi_n = \sum_{k\geqslant n} \exp(\zeta_{k,n}) \geqslant \exp(\zeta_{n,n}) = 1$. Now suppose that $(\xi_n)_{n\in\mathbb{N}}$ is a sequence in $[0,+\infty[$ such that there exists a sequence $(\varepsilon_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ that satisfies

$$\sum_{n\in\mathbb{N}} \chi_n\varepsilon_n < +\infty \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n. \tag{2.46}$$

Set $\theta_0 = 0$ and $(\forall n \in \mathbb{N})$ $\theta_{n+1} = [\xi_{n+1} - \xi_n]^+$ and $\nu_n = \mu_{n,n} - 2$. It results from (2.46), (2.44), and the inequalities $\xi_1 - \xi_0 \leqslant (\mu_{0,0}-1)\xi_0 + \varepsilon_0$ and

$$\begin{aligned}(\forall n \in \mathbb{N}\smallsetminus\{0\}) \quad \xi_{n+1} - \xi_n &\leqslant (\mu_{n,n}-1)\xi_n + (1-\mu_{n,n})\xi_{n-1} + \varepsilon_n \\ &= (\mu_{n,n}-1)(\xi_n - \xi_{n-1}) + \varepsilon_n,\end{aligned} \tag{2.47}$$

that $(\forall n \in \mathbb{N})\ \theta_{n+1} \leqslant (\mu_{n,n} - 1)\theta_n + \varepsilon_n = (1 + \nu_n)\theta_n + \varepsilon_n$. Consequently, we derive from Lemma 2.10 and (2.45) that $(\forall n \in \mathbb{N})\ \theta_{n+1} \leqslant \sum_{k=0}^{n} \varepsilon_k \exp(\zeta_{n,k})$. Using [31, Theorem 141], this yields

$$\sum_{n\in\mathbb{N}} \theta_{n+1} \leqslant \sum_{n\in\mathbb{N}} \sum_{k=0}^{n} \varepsilon_k \exp(\zeta_{n,k}) = \sum_{k\in\mathbb{N}} \varepsilon_k \sum_{n\geqslant k} \exp(\zeta_{n,k}) = \sum_{k\in\mathbb{N}} \varepsilon_k \chi_k. \tag{2.48}$$

Now set $(\forall n \in \mathbb{N})\ \omega_n = \xi_n - \sum_{k=0}^{n} \theta_k$. Since $\sum_{k\in\mathbb{N}} \chi_k \varepsilon_k < +\infty$, we infer from (2.48) that $\sum_{n\in\mathbb{N}} \theta_n < +\infty$. Thus, since $\inf_{n\in\mathbb{N}} \xi_n \geqslant 0$, $(\omega_n)_{n\in\mathbb{N}}$ is bounded below and

$$(\forall n \in \mathbb{N}) \quad \omega_{n+1} = \xi_{n+1} - \theta_{n+1} - \sum_{k=0}^{n} \theta_k \leqslant \xi_{n+1} - \xi_{n+1} + \xi_n - \sum_{k=0}^{n} \theta_k = \omega_n. \tag{2.49}$$

Altogether, $(\omega_n)_{n\in\mathbb{N}}$ converges, and so does therefore $(\xi_n)_{n\in\mathbb{N}}$. $\square$

2.2.3 Asymptotic behavior of Algorithm 2.8

The main result of the paper is the following theorem, which analyzes the asymptotic behavior of Algorithm 2.8.

Theorem 2.15 *Consider the setting of Algorithm 2.8. For every $n \in \mathbb{N}$, define*

$$\vartheta_n = \lambda_n \sum_{i=1}^{m} \|e_{i,n}\| \quad \textit{and} \quad (\forall i \in \{1, \ldots, m\}) \quad T_{i+,n} = \begin{cases} T_{i+1,n} \cdots T_{m,n}, & \textit{if } i \neq m; \\ \mathrm{Id}\,, & \textit{if } i = m, \end{cases} \tag{2.50}$$

and set

$$\nu_n \colon S \to [0, +\infty[\, : x \mapsto \vartheta_n \big(2\|\overline{x}_n - x\| + \vartheta_n\big). \tag{2.51}$$

Then the following hold :

(i) *Let $n \in \mathbb{N}$ and $x \in S$. Then $\|x_{n+1} - x\| \leqslant \sum_{j=0}^{n} |\mu_{n,j}|\, \|x_j - x\| + \vartheta_n$.*

(ii) *Let $n \in \mathbb{N}$ and $x \in S$. Then*

$$\begin{aligned}\|x_{n+1} - x\|^2 \leqslant{} & \sum_{j=0}^{n} \mu_{n,j} \|x_j - x\|^2 - \frac{1}{2} \sum_{j=0}^{n} \sum_{k=0}^{n} \mu_{n,j}\mu_{n,k} \|x_j - x_k\|^2 \\ & - \lambda_n (1/\phi_n - \lambda_n) \|T_n \overline{x}_n - \overline{x}_n\|^2 + \nu_n(x).\end{aligned}$$

(iii) *Let $n \in \mathbb{N}$ and $x \in S$. Then*

$$\begin{aligned}\|x_{n+1} - x\|^2 \leqslant{} & \sum_{j=0}^{n} \mu_{n,j} \|x_j - x\|^2 - \frac{1}{2} \sum_{j=0}^{n} \sum_{k=0}^{n} \mu_{n,j}\mu_{n,k} \|x_j - x_k\|^2 \\ & + \lambda_n (\lambda_n - 1) \|T_n \overline{x}_n - \overline{x}_n\|^2 \\ & - \lambda_n \max_{1\leqslant i\leqslant m} \left(\frac{1 - \alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n}) T_{i+,n} \overline{x}_n - (\mathrm{Id} - T_{i,n}) T_{i+,n} x\|^2 \right) + \nu_n(x).\end{aligned}$$

Now assume that, in addition,

$$\sum_{n\in\mathbb{N}}\chi_n\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j-x_k\|^2<+\infty \quad \textit{and} \quad (\forall x\in S)\quad \sum_{n\in\mathbb{N}}\chi_n\nu_n(x)<+\infty. \tag{2.52}$$

Then the following hold :

(iv) *Let $x\in S$. Then $(\|x_n-x\|)_{n\in\mathbb{N}}$ converges.*

(v) $\lambda_n(1/\phi_n-\lambda_n)\|T_n\overline{x}_n-\overline{x}_n\|^2\to 0.$

(vi) $\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^+\|x_j-x_k\|^2\to 0.$

(vii) *Suppose that*

$$(\exists\,\varepsilon\in\,]0,1[)(\forall n\in\mathbb{N})\quad \lambda_n\leqslant(1-\varepsilon)/\phi_n. \tag{2.53}$$

Then $x_{n+1}-\overline{x}_n\to 0$. In addition, if every weak sequential cluster point of $(\overline{x}_n)_{n\in\mathbb{N}}$ is in S, then there exists $x\in S$ such that $x_n\rightharpoonup x$.

(viii) *Suppose that $(\overline{x}_n)_{n\in\mathbb{N}}$ has a strong cluster point x in S and that (2.53) holds. Then $x_n\to x$.*

(ix) *Let $x\in S$ and suppose that $(\exists\,\varepsilon\in\,]0,1[)(\forall n\in\mathbb{N})$ $\lambda_n\leqslant\varepsilon+(1-\varepsilon)/\phi_n$. Then*

$$\lambda_n\max_{1\leqslant i\leqslant m}\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}-T_{i,n})T_{i+,n}\overline{x}_n-(\mathrm{Id}-T_{i,n})T_{i+,n}x\|^2\to 0.$$

Proof. Let $n\in\mathbb{N}$ and set

$$e_n=T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n+e_{m,n})+e_{m-1,n}\cdots\big)+e_{2,n}\Big)+e_{1,n}-T_n\overline{x}_n. \tag{2.54}$$

If $m>1$, using the nonexpansiveness of the operators $(T_{i,n})_{1\leqslant i\leqslant m-1}$, we obtain

$$\begin{aligned}
&\|e_n\|\\
&\leqslant\|e_{1,n}\|+\left\|T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n+e_{m,n})+e_{m-1,n}\cdots\big)+e_{2,n}\Big)-T_{1,n}\cdots T_{m,n}\overline{x}_n\right\|\\
&\leqslant\|e_{1,n}\|+\\
&\quad\left\|T_{2,n}\Big(T_{3,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n+e_{m,n})+e_{m-1,n}\cdots\big)+e_{3,n}\Big)+e_{2,n}-T_{2,n}\cdots T_{m,n}\overline{x}_n\right\|\\
&\leqslant\|e_{1,n}\|+\|e_{2,n}\|+\\
&\quad\left\|T_{3,n}\Big(T_{4,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n+e_{m,n})+e_{m-1,n}\cdots\big)+e_{4,n}\Big)+e_{3,n}-T_{3,n}\cdots T_{m,n}\overline{x}_n\right\|\\
&\ \vdots\\
&\leqslant\sum_{i=1}^{m}\|e_{i,n}\|.
\end{aligned} \tag{2.55}$$

Thus, we infer from (2.50) that

$$\lambda_n\|e_n\| \leqslant \vartheta_n. \tag{2.56}$$

On the other hand, we derive from (2.26) and (2.54) that

$$x_{n+1} = \overline{x}_n + \lambda_n\big(T_n\overline{x}_n + e_n - \overline{x}_n\big). \tag{2.57}$$

Now set

$$R_n = \frac{1}{\phi_n}T_n - \frac{1-\phi_n}{\phi_n}\,\mathrm{Id} \quad \text{and} \quad \eta_n = \lambda_n\phi_n. \tag{2.58}$$

Then $\eta_n \in \left]0,1\right]$, $\operatorname{Fix} R_n = \operatorname{Fix} T_n$, and R_n is quasinonexpansive since T_n is averaged quasinonexpansive with constant ϕ_n by Remark 2.9(i). Furthermore, (2.57) can be written as

$$x_{n+1} = \overline{x}_n + \eta_n\big(R_n\overline{x}_n - \overline{x}_n\big) + \lambda_n e_n. \tag{2.59}$$

Next, we define

$$z_n = \overline{x}_n + \lambda_n(T_n\overline{x}_n - \overline{x}_n) = \overline{x}_n + \eta_n(R_n\overline{x}_n - \overline{x}_n). \tag{2.60}$$

Let $x \in S$. Since $x \in \operatorname{Fix} R_n$ and R_n is quasinonexpansive, we have

$$\begin{aligned}\|z_n - x\| &= \|(1-\eta_n)(\overline{x}_n - x) + \eta_n(R_n\overline{x}_n - x)\|\\ &\leqslant (1-\eta_n)\|\overline{x}_n - x\| + \eta_n\|R_n\overline{x}_n - x\|\\ &\leqslant \|\overline{x}_n - x\|.\end{aligned} \tag{2.61}$$

Hence, (2.59) and (2.56) yield

$$\|x_{n+1} - x\| \leqslant \|z_n - x\| + \lambda_n\|e_n\| \leqslant \|z_n - x\| + \vartheta_n. \tag{2.62}$$

In turn, it follows from (2.61) and (2.51) that

$$\|x_{n+1} - x\|^2 \leqslant \|z_n - x\|^2 + 2\vartheta_n\|z_n - x\| + \vartheta_n^2 \leqslant \|z_n - x\|^2 + \nu_n(x). \tag{2.63}$$

In addition, [10, Lemma 2.14(ii)] yields

$$\|\overline{x}_n - x\|^2 = \Bigg\|\sum_{j=0}^{n}\mu_{n,j}(x_j - x)\Bigg\|^2 = \sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 - \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2. \tag{2.64}$$

(i) : By (2.62) and (2.61),

$$(\forall n \in \mathbb{N})(\forall x \in S) \quad \|x_{n+1} - x\| \leqslant \|\overline{x}_n - x\| + \vartheta_n \leqslant \sum_{j=0}^{n}|\mu_{n,j}|\,\|x_j - x\| + \vartheta_n. \tag{2.65}$$

(ii) : Let $n \in \mathbb{N}$ and $x \in S$. Since

$$\begin{aligned}\|z_n - x\|^2 &= \|(1-\eta_n)(\overline{x}_n - x) + \eta_n(R_n\overline{x}_n - x)\|^2 \\ &= (1-\eta_n)\|\overline{x}_n - x\|^2 + \eta_n\|R_n\overline{x}_n - x\|^2 - \eta_n(1-\eta_n)\|R_n\overline{x}_n - \overline{x}_n\|^2 \\ &\leqslant \|\overline{x}_n - x\|^2 - \eta_n(1-\eta_n)\|R_n\overline{x}_n - \overline{x}_n\|^2, \end{aligned} \tag{2.66}$$

we deduce from (2.63) and (2.58) that

$$\begin{aligned}\|x_{n+1} - x\|^2 &\leqslant \|z_n - x\|^2 + \nu_n(x) \\ &\leqslant \|\overline{x}_n - x\|^2 - \eta_n(1-\eta_n)\|R_n\overline{x}_n - \overline{x}_n\|^2 + \nu_n(x) \\ &= \|\overline{x}_n - x\|^2 - \lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 + \nu_n(x). \end{aligned} \tag{2.67}$$

In view of (2.64), we obtain the announced inequality.

(iii) : Let $n \in \mathbb{N}$ and $x \in S$. We derive from [10, Proposition 4.35] that

$$(\forall i \in \{1,\ldots,m-1\})(\forall (u,v) \in \mathcal{H}^2)$$
$$\|T_{i,n}u - T_{i,n}v\|^2 \leqslant \|u-v\|^2 - \frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}\,-T_{i,n})u - (\mathrm{Id}\,-T_{i,n})v\|^2. \tag{2.68}$$

If $m > 1$, using this inequality successively for $i = 1,\ldots,m-1$ leads to

$$\begin{aligned}\|T_n\overline{x}_n - x\|^2 &= \|T_{1,n}\cdots T_{m,n}\overline{x}_n - T_{1,n}\cdots T_{m,n}x\|^2 \\ &\leqslant \|T_{m,n}\overline{x}_n - T_{m,n}x\|^2 - \sum_{i=1}^{m-1}\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}\,-T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id}\,-T_{i,n})T_{i+,n}x\|^2 \\ &\leqslant \|T_{m,n}\overline{x}_n - T_{m,n}x\|^2 \\ &\quad - \max_{1\leqslant i\leqslant m-1}\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}\,-T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id}\,-T_{i,n})T_{i+,n}x\|^2. \end{aligned} \tag{2.69}$$

Note that, in cases (a) and (c) of Problem 2.7,

$$\|T_{m,n}\overline{x}_n - T_{m,n}x\|^2 \leqslant \|\overline{x}_n - x\|^2 - \frac{1-\alpha_{m,n}}{\alpha_{m,n}}\|(\mathrm{Id}\,-T_{m,n})\overline{x}_n - (\mathrm{Id}\,-T_{m,n})x\|^2. \tag{2.70}$$

This inequality remains valid in case (b) of Problem 2.7 since [10, Proposition 4.49(i)] implies that

$$\mathrm{Fix}\,(T_{1,n}\cdots T_{m,n}) = \bigcap_{i=1}^{m}\mathrm{Fix}\,T_{i,n} \tag{2.71}$$

and, therefore, that $x \in \mathrm{Fix}\,T_{m,n}$. Altogether, we deduce from (2.69) and (2.70) that

$$\|T_n\overline{x}_n - x\|^2 \leqslant \|\overline{x}_n - x\|^2 - \max_{1\leqslant i\leqslant m}\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}\,-T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id}\,-T_{i,n})T_{i+,n}x\|^2. \tag{2.72}$$

Hence, it follows from (2.60) that

$$\begin{aligned}\|z_n - x\|^2 &= \|(1-\lambda_n)(\overline{x}_n - x) + \lambda_n(T_n\overline{x}_n - x)\|^2 \\ &= (1-\lambda_n)\|\overline{x}_n - x\|^2 + \lambda_n\|T_n\overline{x}_n - x\|^2 + \lambda_n(\lambda_n - 1)\|T_n\overline{x}_n - \overline{x}_n\|^2 \\ &\leqslant \|\overline{x}_n - x\|^2 - \lambda_n \max_{1\leqslant i\leqslant m} \frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id}\, - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id}\, - T_{i,n})T_{i+,n}x\|^2 \\ &\quad + \lambda_n(\lambda_n - 1)\|T_n\overline{x}_n - \overline{x}_n\|^2. \end{aligned} \tag{2.73}$$

In view of (2.63) and (2.64), the inequality is established.

(iv) : Let $x \in S$ and set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \xi_n = \|x_n - x\|^2 \\ \varepsilon_n = \nu_n(x) + \dfrac{1}{2}\displaystyle\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j - x_k\|^2. \end{cases} \tag{2.74}$$

Since $\inf_{n\in\mathbb{N}} \lambda_n(1/\phi_n - \lambda_n) \geqslant 0$, (2.52) and (ii) imply that

$$\sum_{n\in\mathbb{N}} \chi_n\varepsilon_n < +\infty \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n. \tag{2.75}$$

In turn, it follows from (2.52) and condition (d) in Algorithm 2.8 that $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges.

(v)–(vi) : Let $x \in S$. Then it follows from (iv) that $\rho = \lim_{n\to+\infty}\|x_n - x\|$ is well defined. Hence, Lemma 2.11 implies that $\sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 \to \rho^2$ and therefore that

$$\sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 - \|x_{n+1} - x\|^2 \to 0. \tag{2.76}$$

Since $\inf_{n\in\mathbb{N}} \chi_n > 0$, (2.52) yields

$$\nu_n(x) \to 0 \quad \text{and} \quad \sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j - x_k\|^2 \to 0. \tag{2.77}$$

It follows from (ii), (2.76), and (2.77) that

$$\begin{aligned} 0 &\leqslant \lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 + \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^+\|x_j - x_k\|^2 \\ &\leqslant \sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 - \|x_{n+1} - x\|^2 + \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j - x_k\|^2 + \nu_n(x) \\ &\to 0, \end{aligned} \tag{2.78}$$

which gives the desired conclusions.

(vii) : Set $\zeta = 1/\varepsilon - 1$. We deduce from (2.52) and (2.51) that $\sum_{n\in\mathbb{N}} \vartheta_n^2 < +\infty$. Hence, it follows from (2.57), (2.56), (2.53), and (v) that

$$\begin{aligned}
\|x_{n+1} - \overline{x}_n\|^2 &\leqslant 2\Big(\lambda_n^2 \big\|T_n\overline{x}_n - \overline{x}_n\big\|^2 + \lambda_n^2\|e_n\|^2\Big) \\
&\leqslant 2\bigg(\frac{\lambda_n}{1/\phi_n - \lambda_n}\lambda_n(1/\phi_n - \lambda_n)\big\|T_n\overline{x}_n - \overline{x}_n\big\|^2 + \vartheta_n^2\bigg) \\
&\leqslant 2\Big(\zeta\lambda_n(1/\phi_n - \lambda_n)\big\|T_n\overline{x}_n - \overline{x}_n\big\|^2 + \vartheta_n^2\Big) \\
&\to 0.
\end{aligned} \tag{2.79}$$

Therefore $x_{n+1} - \overline{x}_n \rightharpoonup 0$ and hence the weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ lie in S. In view of (iv) and [10, Lemma 2.47], the claim is proved.

(viii) : Since $x_{n+1} - \overline{x}_n \to 0$ by (2.79), $(x_n)_{n\in\mathbb{N}}$ has a strong cluster point $x \in S$. In view of (iv), $x_n \to x$.

(ix) : Set $\zeta = 1/\varepsilon - 1$. Then, for every $n \in \mathbb{N}$, $\lambda_n \leqslant 1/(1+\zeta) + \zeta/(\phi_n(1+\zeta))$ and hence $(1+\zeta)\lambda_n - 1 \leqslant \zeta/\phi_n$, i.e., $\lambda_n - 1 \leqslant \zeta(1/\phi_n - \lambda_n)$. We therefore derive from (iii), (2.76), (2.52), (v), and (2.77) that

$$\begin{aligned}
0 &\leqslant \lambda_n \max_{1\leqslant i\leqslant m} \frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id}\, - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id}\, - T_{i,n})T_{i+,n}x\|^2 \\
&\leqslant \sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 - \|x_{n+1} - x\|^2 + \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j - x_k\|^2 \\
&\quad + \lambda_n(\lambda_n - 1)\|T_n\overline{x}_n - \overline{x}_n\|^2 + \nu_n(x) \\
&\leqslant \sum_{j=0}^{n} \mu_{n,j}\|x_j - x\|^2 - \|x_{n+1} - x\|^2 + \frac{1}{2}\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j - x_k\|^2 \\
&\quad + \zeta\lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 + \nu_n(x) \\
&\to 0,
\end{aligned} \tag{2.80}$$

which shows the assertion. ☐

Next, we present two corollaries that are instrumental in the analysis of two important special cases of our framework : mean value and inertial multilayer algorithms.

Corollary 2.16 *Consider the setting of Algorithm 2.8 and define $(\vartheta_n)_{n\in\mathbb{N}}$ as in (2.50). Assume that*

$$\inf_{n\in\mathbb{N}} \min_{0\leqslant j\leqslant n} \mu_{n,j} \geqslant 0 \quad \textit{and} \quad \sum_{n\in\mathbb{N}} \chi_n\vartheta_n < +\infty. \tag{2.81}$$

Then the following hold :

(i) $\sum_{j=0}^{n}\sum_{k=0}^{n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2 \to 0$.

(ii) *Let* $x \in S$ *and suppose that* $(\exists\,\varepsilon \in \left]0,1\right[)(\forall n \in \mathbb{N})\ \lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n$. *Then*

$$\lambda_n \max_{1\leqslant i\leqslant m} \frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2 \to 0.$$

(iii) *Suppose that every weak sequential cluster point of* $(\overline{x}_n)_{n\in\mathbb{N}}$ *is in* S *and that* $(\exists\,\varepsilon \in \left]0,1\right[)(\forall n \in \mathbb{N})\ \lambda_n \leqslant (1-\varepsilon)/\phi_n$. *Then* $x_{n+1} - \overline{x}_n \to 0$ *and there exists* $x \in S$ *such that* $x_n \rightharpoonup x$.

(iv) *Suppose that every weak sequential cluster point of* $(\overline{x}_n)_{n\in\mathbb{N}}$ *is in* S*, that* $\sup_{n\in\mathbb{N}} \phi_n < 1$*, and that* $(\exists\,\varepsilon \in \left]0,1\right[)(\forall n \in \mathbb{N})\ \lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n$. *Then* $x_{n+1} - \overline{x}_n \to 0$ *and there exists* $x \in S$ *such that* $x_n \rightharpoonup x$.

(v) *Suppose that every weak sequential cluster point of* $(\overline{x}_n)_{n\in\mathbb{N}}$ *is in* S *and that* $\inf_{n\in\mathbb{N}} \mu_{n,n} > 0$. *Then* $x_n - \overline{x}_n \to 0$ *and there exists* $x \in S$ *such that* $x_n \rightharpoonup x$.

Proof. We derive from Theorem 2.15(i) that $(\forall n \in \mathbb{N})\ \|x_{n+1} - x\| \leqslant \sum_{j=0}^{n} \mu_{n,j}\|x_j - x\| + \vartheta_n$. In turn, it follows from condition (d) in Algorithm 2.8 that $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges. As a result, $(x_n)_{n\in\mathbb{N}}$ is bounded and (2.81) therefore implies (2.52).

(i)–(iii) : These follow respectively from items (vi), (ix), and (vii) in Theorem 2.15.

(iv) : Set $\delta = \varepsilon(1 - \sup_{n\in\mathbb{N}} \phi_n)$. Then $\delta \in \left]0, \varepsilon\right[$ and $(\forall n \in \mathbb{N})\ (\varepsilon - \delta)/\phi_n \geqslant \varepsilon$. Hence,

$$(\forall n \in \mathbb{N}) \quad \lambda_n \leqslant \varepsilon + \frac{1-\varepsilon}{\phi_n} = \varepsilon + \frac{1-\delta}{\phi_n} - \frac{\varepsilon - \delta}{\phi_n} \leqslant \frac{1-\delta}{\phi_n}. \tag{2.82}$$

The claim therefore follows from (iii).

(v) : Set

$$\theta = \frac{1}{\inf\limits_{n\in\mathbb{N}} \mu_{n,n}} \quad \text{and} \quad (\forall n \in \mathbb{N})(\forall j \in \{0,\ldots,n\}) \quad \gamma_{n,j} = \begin{cases} \dfrac{\mu_{n,n}+1}{2}, & \text{if } j = n; \\ \dfrac{\mu_{n,j}}{2}, & \text{if } j < n. \end{cases} \tag{2.83}$$

Then, using Apollonius' identity, [10, Lemma 2.12(iv)], and (2.83), we obtain

$$
\begin{aligned}
\frac{1}{4}\|\overline{x}_n - x_n\|^2 &= \frac{1}{2}\Big(\|\overline{x}_n - x\|^2 + \|x_n - x\|^2\Big) - \left\|\frac{\overline{x}_n + x_n}{2} - x\right\|^2 \\
&= \frac{1}{2}\Big(\|\overline{x}_n - x\|^2 + \|x_n - x\|^2\Big) - \left\|\sum_{j=0}^{n}\gamma_{n,j}(x_j - x)\right\|^2 \\
&= \frac{1}{2}\Big(\|\overline{x}_n - x\|^2 + \|x_n - x\|^2\Big) - \sum_{j=0}^{n}\gamma_{n,j}\|x_j - x\|^2 + \sum_{0\leqslant j<k\leqslant n}\gamma_{n,j}\gamma_{n,k}\|x_j - x_k\|^2 \\
&\leqslant \frac{1}{2}\left(\sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 + \|x_n - x\|^2\right) - \sum_{j=0}^{n}\gamma_{n,j}\|x_j - x\|^2 \\
&\quad + \frac{1}{4}\left(\sum_{0\leqslant j<k<n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2 + \sum_{j=0}^{n-1}\mu_{n,j}(\mu_{n,n}+1)\|x_j - x_n\|^2\right) \\
&\leqslant \frac{1}{2}\left(\sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 + \|x_n - x\|^2\right) - \sum_{j=0}^{n}\gamma_{n,j}\|x_j - x\|^2 \\
&\quad + \frac{1}{4}\left(\sum_{0\leqslant j<k\leqslant n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2 + \theta\sum_{j=0}^{n-1}\mu_{n,j}\mu_{n,n}\|x_j - x_n\|^2\right) \\
&\leqslant \frac{1}{2}\left(\sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 + \|x_n - x\|^2\right) - \sum_{j=0}^{n}\gamma_{n,j}\|x_j - x\|^2 \\
&\quad + \frac{1+\theta}{4}\sum_{0\leqslant j<k\leqslant n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2. \qquad (2.84)
\end{aligned}
$$

Next, let us set $\rho = \lim\|x_n - x\|^2$. Then it follows from Lemma 2.11 that $\sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 \to \rho$ and $\sum_{j=0}^{n}\gamma_{n,j}\|x_j - x\|^2 \to \rho$. On the other hand, (i) asserts that $\sum_{0\leqslant j<k\leqslant n}\mu_{n,j}\mu_{n,k}\|x_j - x_k\|^2 \to 0$. Altogether, (2.84) yields $\|\overline{x}_n - x_n\| \to 0$. Thus, the weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ belong to S, and the conclusion follows from the fact that $(\|x_n - x\|)_{n\in\mathbb{N}}$ converges and from [10, Lemma 2.47]. □

Corollary 2.17 *Consider the setting of Algorithm 2.8 with* $(\forall i \in \{1,\ldots,m\})(\forall n \in \mathbb{N})$ $e_{i,n} = 0$. *Set* $x_{-1} = x_0$ *and suppose that there exists a sequence* $(\eta_n)_{n\in\mathbb{N}}$ *in* $[0,1[$ *such that* $\eta_0 = 0$ *and*

$$
(\forall n \in \mathbb{N})(\forall j \in \{0,\ldots,n\}) \quad \mu_{n,j} = \begin{cases} 1+\eta_n, & \text{if } j = n;\\ -\eta_n, & \text{if } j = n-1;\\ 0, & \text{if } j < n-1. \end{cases} \qquad (2.85)
$$

For every $n \in \mathbb{N}$, set

$$(\forall k \in \mathbb{N}) \quad \zeta_{k,n} = \begin{cases} 0, & \text{if } k \leqslant n; \\ \displaystyle\sum_{j=n+1}^{k} (\eta_j - 1), & \text{if } k > n, \end{cases} \tag{2.86}$$

and assume that $\chi_n = \sum_{k \geqslant n} \exp(\zeta_{k,n})$. Suppose that one of the following is satisfied :

(a) $\sum_{n\in\mathbb{N}} \chi_n \eta_n \|x_n - x_{n-1}\|^2 < +\infty$.

(b) $\sum_{n\in\mathbb{N}} n\|x_n - x_{n-1}\|^2 < +\infty$ *and there exists $\tau \in [2, +\infty[$ such that $(\forall n \in \mathbb{N} \smallsetminus \{0\})$ $\eta_n = (n-1)/(n+\tau)$.*

(c) $\sum_{n\in\mathbb{N}} \eta_n \|x_n - x_{n-1}\|^2 < +\infty$ *and there exists $\eta \in [0, 1[$ such that $(\forall n \in \mathbb{N})$ $\eta_n \leqslant \eta$.*

(d) *Set $(\forall n \in \mathbb{N})$ $\omega_n = 1/\phi_n - \lambda_n$. There exist $(\sigma, \vartheta) \in\,]0, +\infty[^2$ and $\eta \in\,]0, 1[$ such that*

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \eta_n \leqslant \eta_{n+1} \leqslant \eta \\ \lambda_n \leqslant \dfrac{\vartheta/\phi_n - \eta\big(\eta(1+\eta) + \eta\vartheta\omega_{n+1} + \sigma\big)}{\vartheta\big(1 + \eta(1+\eta) + \eta\vartheta\omega_{n+1} + \sigma\big)} \\ \dfrac{\eta^2(1+\eta) + \eta\sigma}{\vartheta} < \dfrac{1}{\phi_n} - \eta^2\omega_{n+1}. \end{cases} \tag{2.87}$$

Then the following hold :

(i) $\lambda_n(1/\phi_n - \lambda_n)\|T_n\overline{x}_n - \overline{x}_n\|^2 \to 0$.

(ii) *Let $x \in S$ and suppose that $(\exists\, \varepsilon \in\,]0, 1[)(\forall n \in \mathbb{N})$ $\lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n$. Then*

$$\lambda_n \max_{1\leqslant i\leqslant m} \frac{1-\alpha_{i,n}}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}\overline{x}_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2 \to 0.$$

(iii) $\overline{x}_n - x_n \to 0$.

(iv) *Suppose that every weak sequential cluster point of $(\overline{x}_n)_{n\in\mathbb{N}}$ is in S. Then there exists $x \in S$ such that $x_n \rightharpoonup x$.*

Proof. In view of Example 2.14, $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ satisfies conditions (a)–(d) in Algorithm 2.8.

(a) : Set $\chi = \inf_{n\in\mathbb{N}} \chi_n$ and define $(\nu_n)_{n\in\mathbb{N}}$ as in (2.51). We have $\sup_{n\in\mathbb{N}}(1+\eta_n) \leqslant 2$ and

$$(\forall n \in \mathbb{N}) \quad \begin{cases} \chi_n \sum_{j=0}^{n} \sum_{k=0}^{n} [\mu_{n,j}\mu_{n,k}]^- \|x_j - x_k\|^2 = (1+\eta_n)\chi_n\eta_n\|x_n - x_{n-1}\|^2 \\ (\forall x \in S) \quad \chi_n\nu_n(x) = 0. \end{cases} \tag{2.88}$$

Hence (2.52) holds, and (i) and (ii) follow from Theorem 2.15(v)&(ix), respectively. Furthermore, (2.85) implies that

$$\|\overline{x}_n - x_n\|^2 \leqslant \eta_n^2\|x_n - x_{n-1}\|^2 \leqslant \eta_n\|x_n - x_{n-1}\|^2 \leqslant \frac{\chi_n\eta_n}{\chi}\|x_n - x_{n-1}\|^2 \to 0. \tag{2.89}$$

Thus, (iii) holds. In turn, the weak sequential cluster points of $(x_n)_{n\in\mathbb{N}}$ belong to S and (iv) therefore follows from Theorem 2.15(iv) and [10, Lemma 2.47].

(b)⇒(a) : It follows from Lemma 2.13(i) that

$$\sum_{n\in\mathbb{N}} \chi_n\eta_n\|x_n - x_{n-1}\|^2 \leqslant \sum_{n\in\mathbb{N}} \frac{n+7}{2}\|x_n - x_{n-1}\|^2 < +\infty. \tag{2.90}$$

(c)⇒(a) : Lemma 2.13(ii) asserts that $\sup_{n\in\mathbb{N}} \chi_n \leqslant e/(1-\eta)$.

(d)⇒(c) : Let $x \in S$. It follows from Theorem 2.15(ii) that

$$\begin{aligned}(\forall n\in\mathbb{N})\quad \|x_{n+1}-x\|^2 \leqslant &(1+\eta_n)\|x_n - x\|^2 - \eta_n\|x_{n-1}-x\|^2 + \eta_n(1+\eta_n)\|x_n - x_{n-1}\|^2\\ &- \lambda_n\left(\frac{1}{\phi_n} - \lambda_n\right)\|T_n\overline{x}_n - \overline{x}_n\|^2.\end{aligned} \tag{2.91}$$

Now set $\beta_{-1} = \|x_0 - x\|^2$ and

$$(\forall n\in\mathbb{N})\quad \beta_n = \|x_n - x\|^2,\quad \delta_n = \|x_n - x_{n-1}\|^2,\quad \text{and}\quad \rho_n = \frac{1}{\eta_n + \lambda_n\vartheta}. \tag{2.92}$$

Then

$$\begin{aligned}(\forall n\in\mathbb{N})\quad \|T_n\overline{x}_n - \overline{x}_n\|^2 &= \frac{1}{\lambda_n^2}\big\|(x_{n+1}-x_n) + \eta_n(x_{n-1}-x_n)\big\|^2\\ &= \frac{1}{\lambda_n^2}\left(\delta_{n+1} + \eta_n^2\delta_n + \eta_n\left(2\left\langle \sqrt{\rho_n}(x_{n+1}-x_n)\;\middle|\; \frac{x_{n-1}-x_n}{\sqrt{\rho_n}}\right\rangle\right)\right)\\ &\geqslant \frac{1}{\lambda_n^2}\left(\delta_{n+1} + \eta_n^2\delta_n - \eta_n\left(\rho_n\delta_{n+1} + \frac{\delta_n}{\rho_n}\right)\right).\end{aligned} \tag{2.93}$$

Thus, we derive from (2.91) that

$$(\forall n\in\mathbb{N})\quad \beta_{n+1} - \beta_n - \eta_n(\beta_n - \beta_{n-1}) \leqslant \frac{(1/\phi_n - \lambda_n)(\eta_n\rho_n - 1)}{\lambda_n}\delta_{n+1} + \gamma_n\delta_n, \tag{2.94}$$

where

$$(\forall n\in\mathbb{N})\quad \gamma_n = \eta_n(1+\eta_n) + \eta_n\left(\frac{1}{\phi_n} - \lambda_n\right)\frac{1-\rho_n\eta_n}{\rho_n\lambda_n} \geqslant 0. \tag{2.95}$$

However, it follows from (2.92) that $(\forall n \in \mathbb{N})\ \vartheta = (1-\rho_n\eta_n)/(\rho_n\lambda_n)$. Hence, (2.95) yields

$$(\forall n\in\mathbb{N})\quad \gamma_n = \eta_n(1+\eta_n) + \eta_n\left(\frac{1}{\phi_n} - \lambda_n\right)\vartheta \leqslant \eta(1+\eta) + \eta\vartheta\omega_n. \tag{2.96}$$

Thus, by Lemma 2.12, $\sum_{n\in\mathbb{N}} \eta_n\delta_n \leqslant \sum_{n\in\mathbb{N}} \delta_n < +\infty$ and we conclude that (c) is satisfied. □

Remark 2.18 In Corollary 2.17, no errors were allowed in the implementation of the operators. It is however possible to allow errors in multilayer inertial methods in certain scenarios. For instance, suppose that in Corollary 2.17 we make the additional assumptions that $\lambda_n \equiv 1$ and that $\bigcup_{n\in\mathbb{N}} \operatorname{ran} T_{1,n}$ is bounded. At the same time, let us introduce errors of such that $(\forall i \in \{1,\ldots,m\})\ \sum_{n\in\mathbb{N}} \chi_n \|e_{i,n}\| < +\infty$. Note that (2.26) becomes

$$\begin{array}{l} \text{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = (1+\eta_n)x_n - \eta_n x_{n-1} \\ x_{n+1} = T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{x}_n + e_{m,n}) + e_{m-1,n}\cdots\big) + e_{2,n}\Big) + e_{1,n}. \end{array}\right. \end{array} \tag{2.97}$$

Hence, the assumptions imply that $(x_n)_{n\in\mathbb{N}}$ is bounded. In turn, $(\overline{x}_n)_{n\in\mathbb{N}}$ is bounded and it follows from (2.51) that $(\forall x \in S)\ \sum_{n\in\mathbb{N}} \chi_n \nu_n(x) < +\infty$. An inspection of the proof of Corollary 2.17 then reveals immediately that its conclusions under any of assumptions (a)–(c) remain true.

2.2.4 Examples and Applications

In this section we exhibit various existing results as special cases of our framework. Our purpose is not to exploit it to its full capacity but rather to illustrate its potential on simple instances. We first recover the main result of [21] on algorithm (2.18).

Example 2.19 We consider the setting studied in [21]. Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of firmly quasinonexpansive operators from $\mathcal{H}$ to $\mathcal{H}$ such that $S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing$. Then the problem of finding a point in S is a special case of Problem 2.7(c) where we assume that $\alpha_{1,n} \equiv 1/2$. In addition, let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \|e_n\| < +\infty$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup_{n\in\mathbb{N}} \lambda_n < 2$, and let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be an array with entries in $[0,+\infty[$ which satisfies the following :

(a) $(\forall n \in \mathbb{N})\ \sum_{j=0}^{n} \mu_{n,j} = 1$.

(b) $(\forall j \in \mathbb{N})\ \lim\limits_{n\to+\infty} \mu_{n,j} = 0$.

(c) Every sequence $(\xi_n)_{n\in\mathbb{N}}$ in $[0,+\infty[$ such that

$$\Big(\exists\, (\varepsilon_n)_{n\in\mathbb{N}} \in [0,+\infty[^{\mathbb{N}}\Big) \quad \begin{cases} \sum_{n\in\mathbb{N}} \varepsilon_n < +\infty \\ (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n \end{cases} \tag{2.98}$$

converges.

Clearly, conditions (a)–(c) above imply that, in Algorithm 2.8, conditions (a)–(d) are satisfied. Now let $x_0 \in \mathcal{H}$, and define a sequence $(x_n)_{n\in\mathbb{N}}$ by

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \overline{x}_n + \lambda_n\big(T_n\overline{x}_n + e_n - \overline{x}_n\big), \quad \text{where } \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j}x_j, \tag{2.99}$$

which corresponds to a 1-layer instance of (2.26). This mean iteration process was seen in [21] to cover several classical mean iteration methods, as well as memoryless convex feasibility algorithms [18] (see also [13]). The result obtained in [21, Theorem 3.5(i)] on the weak convergence of $(x_n)_{n\in\mathbb{N}}$ to a point in S corresponds to the special case of Corollary 2.16(iii) in which we further set $\chi_n \equiv 1$.

Next, we retrieve the main result of [25] on the convergence of an overrelaxed version of (2.21) and the special cases discussed there, in particular those of [19].

Example 2.20 We consider the setting studied in [25], which corresponds to Problem 2.7(a). Given $x_0 \in \mathcal{H}$ and sequences $(e_{1,n})_{n\in\mathbb{N}}$, $\ldots$, $(e_{m,n})_{n\in\mathbb{N}}$ in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \lambda_n \sum_{i=1}^{m} \|e_{i,n}\| < +\infty$, construct a sequence $(x_n)_{n\in\mathbb{N}}$ via the m-layer recursion

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \\ \lambda_n\Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}x_n + e_{m,n}) + e_{m-1,n}\cdots\big) + e_{2,n}\Big) + e_{1,n} - x_n\Big), \\ \text{where} \quad 0 < \lambda_n \leqslant \varepsilon + (1-\varepsilon)/\phi_n. \tag{2.100}$$

Note that (2.100) corresponds to the memoryless version of (2.26). The result on the weak convergence of $(x_n)_{n\in\mathbb{N}}$ obtained in [25, Theorem 3.5(iii)] corresponds to the special case of Corollary 2.16(iv) in which the following additional assumptions are made :

(a) $(\forall n \in \mathbb{N})\ \chi_n = 1$ and $(\forall j \in \{0,\ldots,n\})\ \mu_{n,j} = \delta_{n,j}$.
(b) $(\exists\, \varepsilon \in \,]0,1[)(\forall n \in \mathbb{N})\ \lambda_n \leqslant (1-\varepsilon)(1/\phi_n + \varepsilon)$.

Note that condition (a) above implies that, in Algorithm 2.8, conditions (a)–(c) trivially hold, while condition (d) follows from [10, Lemma 5.31]. We also observe that [25, Theorem 3.5(iii)] itself extends the results of [19, Section 3], where the relaxation parameters $(\lambda_n)_{n\in\mathbb{N}}$ are confined to $]0,1]$.

The next two examples feature mean value and inertial iterations in the case of a single quasinonexpansive operator. As is easily seen, the memoryless algorithm (2.15) can fail to produce a convergent sequence in this scenario.

Example 2.21 Let $T\colon \mathcal{H} \to \mathcal{H}$ be a quasinonexpansive operator such that $\mathrm{Id} - T$ is demiclosed at 0 and $\operatorname{Fix} T \neq \varnothing$, let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be an array in $[0,+\infty[$ that satisfies conditions (a)–(d) in Algorithm 2.8 with $\chi_n \equiv 1$ and such that $\inf_{n\in\mathbb{N}}\ \mu_{n+1,n}\mu_{n+1,n+1} > 0$, and let $(e_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \|e_n\| < +\infty$. Let $x_0 \in \mathcal{H}$ and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = T\overline{x}_n + e_n, \quad \text{where} \quad \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j. \tag{2.101}$$

Then $T\overline{x}_n - \overline{x}_n \to 0$ and there exists $x \in \operatorname{Fix} T$ such that $x_n \rightharpoonup x$ and $\overline{x}_n \rightharpoonup x$.

Proof. We apply Corollary 2.16 in the setting of Problem 2.7(c) with $T_{1,n} \equiv T$, $\alpha_{1,n} \equiv 1$, $\phi_n \equiv 1$, and $\lambda_n \equiv 1$. First, note that (2.81) is satisfied. Furthermore, Corollary 2.16(v) entails that $\overline{x}_n - x_n \to 0$, while Corollary 2.16(i) yields $\mu_{n+1,n}\mu_{n+1,n+1}\|x_{n+1} - x_n\|^2 \to 0$ and hence $x_{n+1} - x_n \to 0$. Therefore $\overline{x}_n - T\overline{x}_n = (\overline{x}_n - x_n) + (x_n - x_{n+1}) + e_n \to 0$. Since $\operatorname{Id} - T$ is demiclosed at 0, it follows that every weak sequential cluster point of $(\overline{x}_n)_{n\in\mathbb{N}}$ is in $\operatorname{Fix} T$. In view of Corollary 2.16(v), the proof is complete. □

Example 2.22 Let $T\colon \mathcal{H} \to \mathcal{H}$ be a quasinonexpansive operator such that $\operatorname{Id} - T$ is demiclosed at 0 and $\operatorname{Fix} T \neq \varnothing$, and let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1[$ such that $\eta_0 = 0$, $\eta = \sup_{n\in\mathbb{N}} \eta_n < 1$, and $(\forall n \in \mathbb{N})$ $\eta_n \leqslant \eta_{n+1}$. Let $(\sigma, \vartheta) \in \left]0,+\infty\right[^2$ be such that $(\eta^2(1+\eta) + \eta\sigma)/\vartheta < 1 - \eta^2$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,1[$ such that $0 < \inf_{n\in\mathbb{N}} \lambda_n \leqslant \sup_{n\in\mathbb{N}} \lambda_n \leqslant (\vartheta - \eta(\eta(1+\eta) + \eta\vartheta + \sigma))/(\vartheta(1 + \eta(1+\eta) + \eta\vartheta + \sigma))$. Let $x_0 \in \mathcal{H}$, set $x_{-1} = x_0$, and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \overline{x}_n + \lambda_n\big(T\overline{x}_n - \overline{x}_n\big), \quad \text{where} \quad \overline{x}_n = (1+\eta_n)x_n - \eta_n x_{n-1}. \tag{2.102}$$

Then $T\overline{x}_n - \overline{x}_n \to 0$ and there exists $x \in \operatorname{Fix} T$ such that $x_n \rightharpoonup x$. In the case when T is nonexpansive, this result appears in [14, Theorem 5].

Proof. This is an instance of Corollary 2.17(d)(i)&(iv) and Problem 2.7(c) in which $T_{1,n} \equiv T$, $\alpha_{1,n} \equiv 1$, and $\phi_n \equiv 1$. Note that condition (d) in Corollary 2.17 is satisfied since $(\forall n \in \mathbb{N})$ $\omega_n = 1 - \lambda_n < 1$. □

Next, we consider applications to monotone operator splitting. Let us recall basic notions about a set-valued operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ [10]. We denote by $\operatorname{ran} A = \big\{u \in \mathcal{H} \;\big|\; (\exists x \in \mathcal{H})\ u \in Ax\big\}$ the range of A, by $\operatorname{dom} A = \big\{x \in \mathcal{H} \;\big|\; Ax \neq \varnothing\big\}$ the domain of A, by $\operatorname{zer} A = \big\{x \in \mathcal{H} \;\big|\; 0 \in Ax\big\}$ the set of zeros of A, by $\operatorname{gra} A = \big\{(x,u) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$ the graph of A, and by A^{-1} the inverse of A, i.e., the operator with graph $\big\{(u,x) \in \mathcal{H} \times \mathcal{H} \;\big|\; u \in Ax\big\}$. The resolvent of A is $J_A = (\operatorname{Id} + A)^{-1}$ and $s\colon \operatorname{dom} A \to \mathcal{H}$ is a selection of A if $(\forall x \in \operatorname{dom} A)$ $s(x) \in Ax$. Moreover, A is monotone if

$$(\forall (x,u) \in \operatorname{gra} A)(\forall (y,v) \in \operatorname{gra} A) \quad \langle x - y \mid u - v\rangle \geqslant 0, \tag{2.103}$$

and maximally monotone if there exists no monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$. In this case, J_A is a firmly nonexpansive operator defined everywhere on $\mathcal{H}$ and the reflector $R_A = 2J_A - \operatorname{Id}$ is nonexpansive. We denote by $\Gamma_0(\mathcal{H})$ the class of proper lower semicontinuous convex functions from $\mathcal{H}$ to $\left]-\infty,+\infty\right]$. Let $f \in \Gamma_0(\mathcal{H})$. For every $x \in \mathcal{H}$, $f + \|x - \cdot\|^2/2$ possesses a unique minimizer, which is denoted by $\operatorname{prox}_f x$. We have $\operatorname{prox}_f = J_{\partial f}$, where

$$\partial f\colon \mathcal{H} \to 2^{\mathcal{H}}\colon x \mapsto \big\{u \in \mathcal{H} \;\big|\; (\forall y \in \mathcal{H})\ \langle y - x \mid u\rangle + f(x) \leqslant f(y)\big\} \tag{2.104}$$

is the Moreau subdifferential of f. Our convergence results will rest on the following asymptotic principle.

Lemma 2.23 *Let A and B be maximally monotone operators from $\mathcal{H}$ to $2^{\mathcal{H}}$, let $(x_n,u_n)_{n\in\mathbb{N}}$ be a sequence in* gra A*, let $(y_n,v_n)_{n\in\mathbb{N}}$ be a sequence in* gra B*, let $x\in\mathcal{H}$, and let $v\in\mathcal{H}$. Suppose that $x_n \rightharpoonup x$, $v_n \rightharpoonup v$, $x_n-y_n\to 0$, and $u_n+v_n\to 0$. Then the following hold :*

(i) $(x,-v)\in$ gra A *and* $(x,v)\in$ gra B.

(ii) $0\in Ax+Bx$ *and* $0\in -A^{-1}(-v)+B^{-1}v$.

Proof. Apply [10, Proposition 26.5] with $\mathcal{K}=\mathcal{H}$ and $L=\mathrm{Id}$. □

As discussed in [19], many splitting methods can be analyzed within the powerful framework of fixed point methods for averaged operators. The analysis provided in the present paper therefore makes it possible to develop new methods in this framework, for instance mean value or inertial versions of standard splitting methods. We provide two such examples below. First, we consider the Peaceman-Rachford splitting method, which typically does not converge unless strong requirements are imposed on the underlying operators [20]. In the spirit of Mann's work [34], we show that mean iterations induce the convergence of this algorithm.

Proposition 2.24 *Let $A\colon\mathcal{H}\to 2^{\mathcal{H}}$ and $B\colon\mathcal{H}\to 2^{\mathcal{H}}$ be maximally monotone operators such that* zer $(A+B)\neq\varnothing$ *and let $\gamma\in\,]0,+\infty[$. Let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}}\|a_n\|<+\infty$ and $\sum_{n\in\mathbb{N}}\|b_n\|<+\infty$, let $x_0\in\mathcal{H}$, and let $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ be an array in $[0,+\infty[$ that satisfies conditions (a)–(d) in Algorithm 2.8 with $\chi_n\equiv 1$ and such that $\inf_{n\in\mathbb{N}}\ \mu_{n+1,n}\mu_{n+1,n+1}>0$. Iterate*

$$
\begin{array}{l}
\text{for } n=0,1,\ldots\\
\left\lfloor
\begin{array}{l}
\overline{x}_n=\displaystyle\sum_{j=0}^{n}\mu_{n,j}x_j\\
y_n=J_{\gamma B}\overline{x}_n+b_n\\
z_n=J_{\gamma A}(2y_n-\overline{x}_n)+a_n\\
x_{n+1}=\overline{x}_n+2(z_n-y_n).
\end{array}
\right.
\end{array}
\tag{2.105}
$$

Then there exists $x\in$ Fix $R_{\gamma A}R_{\gamma B}$ *such that $x_n\rightharpoonup x$ and $\overline{x}_n\rightharpoonup x$. Now set $y=J_{\gamma B}x$. Then $y\in$* zer $(A+B)$*, $z_n-y_n\to 0$, $y_n\rightharpoonup y$, and $z_n\rightharpoonup y$.*

Proof. Set $T=R_{\gamma A}R_{\gamma B}$ and $(\forall n\in\mathbb{N})$ $e_n=2a_n+R_{\gamma A}(R_{\gamma B}\overline{x}_n+2b_n)-R_{\gamma A}(R_{\gamma B}\overline{x}_n)$. Then T is nonexpansive, $\mathrm{Id}-T$ is therefore demiclosed, and, since zer $(A+B)\neq\varnothing$, [10, Proposition 26.1(iii)(b)] yields Fix $T=$ Fix $R_{\gamma A}R_{\gamma B}\neq\varnothing$. In addition, we derive from (2.105) that

$$
(\forall n\in\mathbb{N})\quad x_{n+1}=T\overline{x}_n+e_n,
\tag{2.106}
$$

where

$$\begin{aligned}\sum_{n\in\mathbb{N}}\|e_n\| &\leqslant \sum_{n\in\mathbb{N}}\big(2\|a_n\| + \|R_{\gamma A}(R_{\gamma B}\overline{x}_n + 2b_n) - R_{\gamma A}(R_{\gamma B}\overline{x}_n)\|\big)\\ &\leqslant \sum_{n\in\mathbb{N}} 2\big(\|a_n\| + \|b_n\|\big)\\ &< +\infty. \end{aligned}\tag{2.107}$$

Consequently, we deduce from Example 2.21 that $(x_n)_{n\in\mathbb{N}}$ and $(\overline{x}_n)_{n\in\mathbb{N}}$ converge weakly to a point $x \in \operatorname{Fix} T = \operatorname{Fix} R_{\gamma A}R_{\gamma B}$, and that $T\overline{x}_n - \overline{x}_n \to 0$. In addition, [10, Proposition 26.1(iii)(b)] asserts that $y \in \operatorname{zer}(A+B)$. Next, we derive from (2.105) and (2.106) that $2(z_n - y_n) = x_{n+1} - \overline{x}_n = (T\overline{x}_n - \overline{x}_n) + e_n \to 0$. It remains to show that $y_n \rightharpoonup y$. Since $(\overline{x}_n)_{n\in\mathbb{N}}$ converges weakly, it is bounded. However, $(\forall n \in \mathbb{N})$ $\|y_n - y_0\| = \|J_{\gamma B}\overline{x}_n - J_{\gamma B}x_0 + b_n\| \leqslant \|\overline{x}_n - x_0\| + \|b_n\|$. Therefore $(y_n)_{n\in\mathbb{N}}$ is bounded. Now let z be a weak sequential cluster point of $(y_n)_{n\in\mathbb{N}}$, say $y_{k_n} \rightharpoonup z$. In view of [10, Lemma 2.46], it is enough to show that $z = y$. To this end, set $(\forall n \in \mathbb{N})$ $v_n = \gamma^{-1}(\overline{x}_n - y_n + b_n)$ and $w_n = \gamma^{-1}(2y_n - \overline{x}_n - z_n + a_n)$. Then $(\forall n \in \mathbb{N})$ $(z_n - a_n, w_n) \in \operatorname{gra} A$ and $(y_n - b_n, v_n) \in \operatorname{gra} B$. In addition, we have $\overline{x}_{k_n} \rightharpoonup x$, $z_{k_n} - a_{k_n} \rightharpoonup z$, $v_{k_n} \rightharpoonup \gamma^{-1}(x - z)$, $(z_{k_n} - a_{k_n}) - (y_{k_n} - b_{k_n}) \to 0$, and $v_{k_n} + w_{k_n} = \gamma^{-1}(y_{k_n} - z_{k_n} + a_{k_n} + b_{k_n}) \to 0$. Hence, we derive from Lemma 2.23(i) that $(z, \gamma^{-1}(x - z)) \in \operatorname{gra} B$, i.e., $z = J_{\gamma B}x = y$. ☐

Remark 2.25 Let f and g be functions in $\Gamma_0(\mathcal{H})$, and specialize Proposition 2.24 to $A = \partial f$ and $B = \partial g$. Then $\operatorname{zer}(A+B) = \operatorname{Argmin}(f+g)$. Moreover, (2.105) becomes

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j}x_j\\ y_n = \operatorname{prox}_{\gamma g}\overline{x}_n + b_n\\ z_n = \operatorname{prox}_{\gamma f}(2y_n - \overline{x}_n) + a_n\\ x_{n+1} = \overline{x}_n + 2(z_n - y_n), \end{array}\right.\end{array}\tag{2.108}$$

and we conclude that there exists a point $y \in \operatorname{Argmin}(f+g)$ such that $y_n \rightharpoonup y$ and $z_n \rightharpoonup y$.

We now propose a new forward-backward splitting framework which includes existing instances as special cases. The following notion will be needed to establish strong convergence properties (see [4, Proposition 2.4] for special cases).

Definition 2.26 [4, Definition 2.3] An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is *demiregular* at $x \in \operatorname{dom} A$ if, for every sequence $(x_n, u_n)_{n\in\mathbb{N}}$ in $\operatorname{gra} A$ and every $u \in Ax$ such that $x_n \rightharpoonup x$ and $u_n \to u$, we have $x_n \to x$.

Proposition 2.27 *Let $\beta \in\,]0, +\infty[$, let $\varepsilon \in\,]0, \min\{1/2, \beta\}[$, let $x_0 \in \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $B\colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive, i.e.,*

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x - y \mid Bx - By\rangle/\beta \geqslant \|Bx - By\|^2. \tag{2.109}$$

Furthermore, let $(\gamma_n)_{n\in\mathbb{N}}$ *be a sequence in* $[\varepsilon, 2\beta/(1+\varepsilon)]$*, let* $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *be a real array that satisfies conditions (a)–(d) in Algorithm 2.8, and let* $(a_n)_{n\in\mathbb{N}}$ *and* $(b_n)_{n\in\mathbb{N}}$ *be sequences in* $\mathcal{H}$ *such that* $\sum_{n\in\mathbb{N}} \chi_n\|a_n\| < +\infty$ *and* $\sum_{n\in\mathbb{N}} \chi_n\|b_n\| < +\infty$*. Suppose that* $\operatorname{zer}(A+B) \neq \varnothing$ *and that*

$$(\forall n\in\mathbb{N}) \quad \lambda_n \in \left[\varepsilon, 1+(1-\varepsilon)\left(1-\frac{\gamma_n}{2\beta}\right)\right], \tag{2.110}$$

and set $(\forall n\in\mathbb{N})$ $\phi_n = 2/(4-\gamma_n/\beta)$*. For every* $n\in\mathbb{N}$*, iterate*

$$x_{n+1} = \overline{x}_n + \lambda_n\Big(J_{\gamma_n A}\big(\overline{x}_n - \gamma_n(B\overline{x}_n + b_n)\big) + a_n - \overline{x}_n\Big), \quad \textit{where} \quad \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j}x_j. \tag{2.111}$$

Suppose that one of the following is satisfied :

(a) $\inf_{n\in\mathbb{N}} \min_{0\leqslant j\leqslant n} \mu_{n,j} \geqslant 0$.

(b) $a_n \equiv b_n \equiv 0$, $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *satisfies (2.85), and one of conditions (a)–(d) in Corollary 2.17 is satisfied.*

Then the following hold :

(i) $J_{\gamma_n A}(\overline{x}_n - \gamma_n B\overline{x}_n) - \overline{x}_n \to 0$.

(ii) *Let* $z\in\operatorname{zer}(A+B)$*. Then* $B\overline{x}_n \to Bz$*.*

(iii) *There exists* $x\in\operatorname{zer}(A+B)$ *such that* $x_n \rightharpoonup x$*.*

(iv) *Suppose that* A *or* B *is demiregular at every point in* $\operatorname{zer}(A+B)$*. Then there exists* $x\in\operatorname{zer}(A+B)$ *such that* $x_n \to x$*.*

Proof. We apply Corollary 2.16 in case (a) and from Corollary 2.17 in case (b). We first note that (2.111) is an instance of Algorithm 2.8 with $m=2$ and $(\forall n\in\mathbb{N})$ $T_{1,n} = J_{\gamma_n A}$, $T_{2,n} = \operatorname{Id} - \gamma_n B$, $e_{1,n} = a_n$, and $e_{2,n} = -\gamma_n b_n$. Indeed, for every $n\in\mathbb{N}$, $T_{1,n}$ is $\alpha_{1,n}$-averaged with $\alpha_{1,n} = 1/2$ [10, Remark 4.34(iii) and Corollary 23.9], $T_{2,n}$ is $\alpha_{2,n}$-averaged with $\alpha_{2,n} = \gamma_n/(2\beta)$ [10, Proposition 4.39], and the averaging constant of $T_{1,n}T_{2,n}$ is therefore given by (2.27) as

$$\frac{\alpha_{1,n}+\alpha_{2,n}-2\alpha_{1,n}\alpha_{2,n}}{1-\alpha_{1,n}\alpha_{2,n}} = \frac{2}{4-\gamma_n/\beta} = \phi_n. \tag{2.112}$$

On the other hand, we are in the setting of Problem 2.7(a) since [10, Proposition 26.1(iv)(a)] yields $(\forall n\in\mathbb{N})$ $\operatorname{Fix}(T_{1,n}T_{2,n}) = \operatorname{zer}(A+B) \neq \varnothing$. We also observe that, in view of (2.110),

$$(\forall n\in\mathbb{N}) \quad \varepsilon \leqslant \lambda_n \leqslant \varepsilon + \frac{1-\varepsilon}{\phi_n}, \quad \frac{1-\alpha_{1,n}}{\alpha_{1,n}} = 1, \quad \text{and} \quad \frac{1-\alpha_{2,n}}{\alpha_{2,n}} \geqslant \varepsilon \tag{2.113}$$

which, by (2.27), yields

$$\sup_{n\in\mathbb{N}} \phi_n \leqslant \frac{1+\varepsilon}{1+2\varepsilon} < 1. \tag{2.114}$$

In addition, it results from (2.112) that $(\forall n \in \mathbb{N})\ \lambda_n \leqslant 1/\phi_n + \varepsilon \leqslant 2 - \gamma_n/(2\beta) + \varepsilon \leqslant 2 + \varepsilon$. Therefore,

$$\begin{cases} \sum_{n\in\mathbb{N}} \chi_n \lambda_n \|e_{1,n}\| = (2+\varepsilon) \sum_{n\in\mathbb{N}} \chi_n \|a_n\| < +\infty \\ \sum_{n\in\mathbb{N}} \chi_n \lambda_n \|e_{2,n}\| \leqslant 2\beta(2+\varepsilon) \sum_{n\in\mathbb{N}} \chi_n \|b_n\| < +\infty, \end{cases} \tag{2.115}$$

which establishes (2.81) for case (a). Altogether, (2.113), Corollary 2.16(ii), and Corollary 2.17(ii) imply that, for every $z \in \operatorname{zer}(A+B)$,

$$\begin{cases} (T_{1,n} - \operatorname{Id})T_{2,n}\overline{x}_n + (T_{2,n} - \operatorname{Id})z = (T_{1,n} - \operatorname{Id})T_{2,n}\overline{x}_n - (T_{1,n} - \operatorname{Id})T_{2,n}z \to 0 \\ (T_{2,n} - \operatorname{Id})\overline{x}_n - (T_{2,n} - \operatorname{Id})z \to 0. \end{cases} \tag{2.116}$$

Now set

$$(\forall n \in \mathbb{N}) \quad y_n = J_{\gamma_n A}(\overline{x}_n - \gamma_n B\overline{x}_n), \quad u_n = \frac{\overline{x}_n - y_n}{\gamma_n} - B\overline{x}_n, \quad \text{and} \quad v_n = B\overline{x}_n, \tag{2.117}$$

and note that

$$(\forall n \in \mathbb{N}) \quad u_n \in Ay_n. \tag{2.118}$$

(i) : Let $z \in \operatorname{zer}(A+B)$. Then (2.116) yields $J_{\gamma_n A}(\overline{x}_n - \gamma_n B\overline{x}_n) - \overline{x}_n = (T_{1,n} - \operatorname{Id})T_{2,n}\overline{x}_n + (T_{2,n} - \operatorname{Id})z + (T_{2,n} - \operatorname{Id})\overline{x}_n - (T_{2,n} - \operatorname{Id})z \to 0$.

(ii) : We derive from (2.116) that

$$\|B\overline{x}_n - Bz\| = \gamma_n^{-1}\|T_{2,n}\overline{x}_n - \overline{x}_n - T_{2,n}z + z\| \leqslant \varepsilon^{-1}\|T_{2,n}\overline{x}_n - \overline{x}_n - T_{2,n}z + z\| \to 0. \tag{2.119}$$

(iii) : Let $(k_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ and let $y \in \mathcal{H}$ be such that $\overline{x}_{k_n} \rightharpoonup y$. In view of Corollary 2.16(iv) in case (a), and Corollary 2.17(iv) in case (b), it remains to show that $y \in \operatorname{zer}(A+B)$. We derive from (i) that $y_n - \overline{x}_n \to 0$. Hence $y_{k_n} \rightharpoonup y$. Now let $z \in \operatorname{zer}(A+B)$. Then (ii) implies that $B\overline{x}_n \to Bz$. Altogether, $y_{k_n} \rightharpoonup y$, $v_{k_n} \rightharpoonup Bz$, $y_{k_n} - \overline{x}_{k_n} \to 0$, $u_{k_n} + v_{k_n} \to 0$, and, for every $n \in \mathbb{N}$, $u_{k_n} \in Ay_{k_n}$ and $v_{k_n} \in B\overline{x}_{k_n}$. It therefore follows from Lemma 2.23(ii) that $y \in \operatorname{zer}(A+B)$.

(iv) : By (iii), there exists $x \in \operatorname{zer}(A+B)$ such that $x_n \rightharpoonup x$. In addition, we derive from Corollary 2.16(iv) or Corollary 2.17(iii) that

$$\overline{x}_n - x_{n+1} \to 0 \quad \text{or} \quad \overline{x}_n - x_n \to 0. \tag{2.120}$$

Hence it follows from (2.117) and (i) that $y_n \rightharpoonup x$, and, from (2.117) and (ii), that $u_n \to -Bx$. In turn, if A is demiregular on $\operatorname{zer}(A+B)$, we derive from (2.118) that $y_n \to x$. Since $y_n - \overline{x}_n \to 0$, (2.120) yields $x_n \to x$. Now suppose that B is demiregular on $\operatorname{zer}(A+B)$. Since $\overline{x}_n \rightharpoonup x$, (ii) implies that $\overline{x}_n \to x$ and it follows from (2.120) that $x_n \to x$. ☐

Remark 2.28 As noted in Remark 2.18, we can allow errors in inertial multilayer methods and, in particular, in the inertial forward-backward algorithm. Thus, suppose that, in Proposition 2.27, $\lambda_n \equiv 1$ and A has bounded domain. Then $\bigcup_{n\in\mathbb{N}} \operatorname{ran} T_{1,n} = \bigcup_{n\in\mathbb{N}} \operatorname{ran}(\operatorname{Id} + \gamma_n A)^{-1} = \operatorname{dom} A$ is bounded. Hence, it follows from Remark 2.18 that, if $\sum_{n\in\mathbb{N}} \chi_n\|a_n\| < +\infty$ and $\sum_{n\in\mathbb{N}} \chi_n\|b_n\| < +\infty$, the conclusions of Proposition 2.27(b) under any of assumptions (a)–(c) of Corollary 2.17 remain true for the inertial forward-backward algorithm

$$\begin{array}{l}\text{for } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} \overline{x}_n = (1+\eta_n)x_n - \eta_n x_{n-1}\\ x_{n+1} = J_{\gamma_n A}\big(\overline{x}_n - \gamma_n(B\overline{x}_n + b_n)\big) + a_n. \end{array}\right.\end{array} \tag{2.121}$$

Remark 2.29 Let $f \in \Gamma_0(\mathcal{H})$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a $1/\beta$-Lipschitzian gradient, and suppose that $\operatorname{Argmin}(f+g) \neq \varnothing$. Then ∇g is β-cocoercive [10, Corollary 18.17]. Upon setting $A = \partial f$ and $B = \nabla g$ in Proposition 2.27, we see that, for every $n \in \mathbb{N}$, (2.111) becomes

$$x_{n+1} = \overline{x}_n + \lambda_n\Big(\operatorname{prox}_{\gamma_n f}\big(\overline{x}_n - \gamma_n(\nabla g(\overline{x}_n) + b_n)\big) + a_n - \overline{x}_n\Big), \quad \text{where} \quad \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j, \tag{2.122}$$

and we conclude that there exists $x \in \operatorname{Argmin}(f+g)$ such that $x_n \rightharpoonup x$ and $\nabla g(\overline{x}_n) \to \nabla g(x)$.

Remark 2.30 Various results on the convergence of the forward-backward splitting algorithm can be recovered from Proposition 2.27.

(i) Suppose that $(\forall n \in \mathbb{N})\ \lambda_n \leqslant 1$ and $(\forall j \in \{0,\ldots,n\})\ \mu_{n,j} = \delta_{n,j}$. Then conditions (a)–(d) in Algorithm 2.8 hold with $\chi_n \equiv 1$ by [10, Lemma 5.31]. In turn, Proposition 2.27(a)(iii) reduces to [19, Corollary 6.5]. In the context of Remark 2.29, Proposition 2.27(a)(iii) captures [24, Theorem 3.4(i)]. In this setting, under suitable conditions on the errors, it is shown in [23, Theorem 3(vi)] that $(f+g)(x_n) - \min(f+g)(\mathcal{H}) = o(1/n)$.

(ii) In the context of Remark 2.29, let $(\eta_n)_{n\in\mathbb{N}}$ be a sequence in $[0,1[$ that satisfies condition (b) in Corollary 2.17, and set $x_{-1} = x_0$ and $\eta_0 = 0$. If $\gamma_n \equiv \gamma_0 \leqslant \beta$ and $\lambda_n \equiv 1$, Proposition 2.27(b)(iii) covers the scheme

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma_0 f}\Big(x_n + \eta_n(x_n - x_{n-1}) - \gamma_0\nabla g\big(x_n + \eta_n(x_n - x_{n-1})\big)\Big) \tag{2.123}$$

studied in [17, Theorem 4.1], where it was established that $\sum_{n\in\mathbb{N}} n\|x_n - x_{n-1}\|^2 < +\infty$. In this case, it is shown in [6, Theorem 1] that $(f+g)(x_n) - \min(f+g)(\mathcal{H}) = o(1/n^2)$.

(iii) If $\lambda_n \equiv 1$, then Proposition 2.27(b)(iii) under hypothesis (c) of Corollary 2.17 establishes a statement made in [33, Theorem 1]. Let us note, however, that the proof of [33] is not convincing as the authors appear to use the weak continuity of some operators which are merely strongly continuous.

(iv) Suppose that $(\forall n \in \mathbb{N})$ $\lambda_n \leqslant (1-\varepsilon)(2+\varepsilon-\gamma_n/(2\beta))$ and $(\forall j \in \{0,\ldots,n\})$ $\mu_{n,j} = \delta_{n,j}$. Then items (a)(iii) and (a)(iv) of Proposition 2.27 capture respectively items (iii) and (iv)(a)&(b) of [25, Proposition 4.4]. In addition, in the context of Remark 2.29, Proposition 2.27(a)(iii) captures [25, Proposition 4.7(iii)].

(v) Proposition 2.27 also applies to the proximal point algorithm. Indeed, when $B = 0$, it suffices to allow $\beta = +\infty$ and $\alpha_{2,n} \equiv 0$, to set $1/{+\infty} = 0$ and $1/0 = +\infty$, and to take $(\gamma_n)_{n\in\mathbb{N}}$ in $[\varepsilon, +\infty[$. In this setting, the proof remains valid and :

(a) Proposition 2.27(b)(i)&(iii) under hypothesis (c) of Corollary 2.17 capture the error-free case of [2, Theorem 3.1], while Theorem 2.15 covers its general case.

(b) Let $\eta \in \left]0, 1/3\right[$, set $\sigma = (1-3\eta)/2$ and $\vartheta = 2/3$, and suppose that $\lambda_n \equiv 1$. Then Proposition 2.27(b) under hypothesis (d) of Corollary 2.17 yields [3, Proposition 2.1].

Next, we derive from Corollary 2.16 a mean value extension of Polyak's subgradient projection method [40] (likewise, an inertial version can be derived from Corollary 2.17).

Example 2.31 Let C be a nonempty closed convex subset of $\mathcal{H}$ with projector P_C, let $f\colon \mathcal{H} \to \mathbb{R}$ be a continuous convex function such that $\operatorname{Argmin}_C f \neq \varnothing$ and $\theta = \min f(C)$ is known. Suppose that one of the following holds :

(i) f is bounded on every bounded subset of $\mathcal{H}$.

(ii) The conjugate f^* of f is supercoercive, i.e., $\lim_{\|u\|\to+\infty} f^*(u)/\|u\| = +\infty$.

(iii) $\mathcal{H}$ is finite-dimensional.

Let $\eta \in \left]0, 1\right[$, let $\varepsilon \in \left]0, \eta/(2+\eta)\right[$, let $(\mu_{n,j})_{n\in\mathbb{N}, 0\leqslant j\leqslant n}$ be an array in $\left[0, +\infty\right[$ that satisfies conditions (a)–(d) in Algorithm 2.8, let $(\xi_n)_{n\in\mathbb{N}}$ be in $[\eta, 2-\eta]$, let $(\lambda_n)_{n\in\mathbb{N}}$ be in $[\varepsilon, (1-\varepsilon)(2-\xi_n/2)]$, let s be a selection of ∂f, and let $x_0 \in C$. Iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
\overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\
x_{n+1} = \begin{cases} \overline{x}_n + \lambda_n \left(P_C\left(\overline{x}_n + \xi_n \dfrac{\theta - f(\overline{x}_n)}{\|s(\overline{x}_n)\|^2} s(\overline{x}_n) \right) - \overline{x}_n \right), & \text{if } s(\overline{x}_n) \neq 0; \\ \overline{x}_n, & \text{if } s(\overline{x}_n) = 0. \end{cases}
\end{array}
\right.
\end{array}
\tag{2.124}
$$

Then there exists $x \in \operatorname{Argmin}_C f$ such that $x_n \rightharpoonup x$.

Proof. Let G be the subgradient projector onto $D = \big\{x \in \mathcal{H} \;\big|\; f(x) \leqslant \theta\big\}$ associated with s, that is,

$$
G\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \begin{cases} x + \dfrac{\theta - f(x)}{\|s(x)\|^2} s(x), & \text{if } f(x) > \theta; \\ x, & \text{if } f(x) \leqslant \theta. \end{cases}
\tag{2.125}
$$

Then G is firmly quasinonexpansive [10, Proposition 29.41(iii)]. Now set $(\forall n \in \mathbb{N})$ $T_{1,n} = P_C$, $T_{2,n} = \mathrm{Id} + \xi_n(G - \mathrm{Id})$, $\alpha_{1,n} = 1/2$, and $\alpha_{2,n} = \xi_n/2$. Then, for every $n \in \mathcal{H}$, $T_{1,n}$ is an $\alpha_{1,n}$-averaged nonexpansive operator [10, Proposition 4.16], $T_{2,n}$ is an $\alpha_{2,n}$-averaged quasinonexpansive operator, [10, Proposition 4.49(i)] yields

$$\operatorname{Fix} T_{1,n}T_{2,n} = \operatorname{Fix} T_{1,n} \cap \operatorname{Fix} T_{2,n} = C \cap D = \operatorname{Argmin}_C f, \tag{2.126}$$

and Remark 2.9(i) asserts that $T_{1,n}T_{2,n}$ is an averaged quasinonexpansive operator with constant $\phi_n = 2/(4 - \xi_n)$ and $\lambda_n \leqslant (1 - \varepsilon)/\phi_n$. Thus, the problem of minimizing f over C is a special case of Problem 2.7(b) with $m = 2$, and (2.124) is a special case of Algorithm 2.8 with $e_{1,n} \equiv 0$ and $e_{2,n} \equiv 0$. Now let $(k_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ and let $x \in \mathcal{H}$ be such that $\overline{x}_{k_n} \rightharpoonup x$. Then, by Corollary 2.16(iii), it remains to show that $x \in C \cap D$. We derive from Corollary 2.16(ii) and (2.126) that $T_{2,k_n}\overline{x}_{k_n} - P_C T_{2,k_n}\overline{x}_{k_n} \to 0$ and $\overline{x}_{k_n} - T_{2,k_n}\overline{x}_{k_n} \to 0$. Therefore $C \ni P_C T_{2,k_n}\overline{x}_{k_n} = (P_C T_{2,k_n}\overline{x}_{k_n} - T_{2,n}\overline{x}_{k_n}) + (T_{2,k_n}\overline{x}_{k_n} - \overline{x}_{k_n}) + \overline{x}_{k_n} \rightharpoonup x$ and, since C is weakly closed, $x \in C$. On the other hand, $\|G\overline{x}_n - \overline{x}_n\| = \|T_{2,n}\overline{x}_n - \overline{x}_n\|/\xi_n \leqslant \|T_{2,n}\overline{x}_n - \overline{x}_n\|/\eta \to 0$. Since (iii)⇒(ii)⇔(i) [10, Proposition 16.20] and (i) imply that $\mathrm{Id} - G$ is demiclosed at 0 [10, Proposition 29.41(vii)], we conclude that $x \in \operatorname{Fix} G = D$. ☐

Remark 2.32 Example 2.31 reverts to Polyak's classical result [40, Theorem 1] in the case when $(\forall n \in \mathbb{N})$ $\lambda_n = 1$ and $(\forall j \in \{0, \ldots, n\})$ $\mu_{n,j} = \delta_{n,j}$. The unrelaxed pattern $\lambda_n \equiv 1$ is indeed achievable because $(\forall n \in \mathbb{N})$ $\lambda_n \in [\varepsilon, (1 - \varepsilon)(2 - \xi_n/2)]$ and $(1 - \varepsilon)(2 - \xi_n/2) \geqslant (1 - \varepsilon)(2 - (2 - \eta)/2) > (1 - \eta/(2 + \eta))(1 + \eta/2) = 1$.

2.3 Bibliographie

[1] F. Alvarez, *On the minimizing property of a second order dissipative system in Hilbert spaces,* SIAM J. Control Optim., 38 (2000), pp. 1102–1119.

[2] F. Alvarez, *Weak convergence of a relaxed and inertial hybrid projection-proximal point algorithm for maximal monotone operators in Hilbert space,* SIAM J. Optim., 14 (2004), pp. 773–782.

[3] F. Alvarez and H. Attouch, *An inertial proximal method for maximal monotone operators via discretization of a nonlinear oscillator with damping,* Set-Valued Anal., 9 (2001), pp. 3–11.

[4] H. Attouch, L. M. Briceño-Arias, and P. L. Combettes, *A parallel splitting method for coupled monotone inclusions,* SIAM J. Control Optim., 48 (2010), pp. 3246–3270.

[5] H. Attouch, X. Goudou, and P. Redont, *The heavy ball with friction method, I. The continuous dynamical system : global exploration of the local minima of real-valued function by asymptotic analysis of a dissipative dynamical system,* Commun. Contemp. Math., 2 (2000), pp. 1–34.

[6] H. Attouch and J. Peypouquet, *The rate of convergence of Nesterov's accelerated forward-backward method is actually faster than $1/k^2$,* SIAM J. Optim., 26 (2016), pp. 1824–1834.

[7] J. B. BAILLON, R. E. BRUCK, AND S. REICH, *On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces,* Houston J. Math., 4 (1978), pp. 1–9.

[8] D. BARRO AND E. CANESTRELLI, *Combining stochastic programming and optimal control to decompose multistage stochastic optimization problems,* OR Spectrum, 38 (2016), pp. 711–742.

[9] H. H. BAUSCHKE AND J. M. BORWEIN, *On projection algorithms for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.

[10] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed., Springer, New York, 2017.

[11] E. BOREL, *Leçons sur les Séries Divergentes,* Gauthier-Villars, Paris, 1901.

[12] D. BORWEIN AND J. BORWEIN, *Fixed point iterations for real functions,* J. Math. Anal. Appl., 157 (1991), pp. 112–126.

[13] J. M. BORWEIN, G. LI, AND M. K. TAM, *Convergence rate analysis for averaged fixed point iterations in common fixed point problems,* SIAM J. Optim., 27 (2017), pp. 1–33.

[14] R. I. BOŢ, E. R. CSETNEK, AND C. HENDRICH, *Inertial Douglas-Rachford splitting for monotone inclusion problems,* Appl. Math. Comput., 256 (2015), pp. 472–487.

[15] C. BREZINSKI AND J.-P. CHEHAB, *Nonlinear hybrid procedures and fixed point iterations,* Numer. Funct. Anal. Optim., 19 (1998), pp. 465–487.

[16] C. L. BYRNE, *Iterative Optimization in Inverse Problems,* CRC Press, Boca Raton, FL, 2014.

[17] A. CHAMBOLLE AND C. DOSSAL, *On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm",* J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[18] P. L. COMBETTES, *Quasi-Fejérian analysis of some optimization algorithms,* in : Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds., Elsevier, New York, 2001, pp. 115–152.

[19] P. L. COMBETTES, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), pp. 475–504.

[20] P. L. COMBETTES, *Iterative construction of the resolvent of a sum of maximal monotone operators,* J. Convex Anal., 16 (2009), pp. 727–748.

[21] P. L. COMBETTES AND T. PENNANEN, *Generalized Mann iterates for constructing fixed points in Hilbert spaces,* J. Math. Anal. Appl., 275 (2002), pp. 521–536.

[22] P. L. COMBETTES AND J.-C. PESQUET, *Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping,* SIAM J. Optim., 25 (2015), pp. 1221–1248.

[23] P. L. COMBETTES, S. SALZO, AND S. VILLA, *Consistent learning by composite proximal thresholding,* Math. Program., published online 2017-03-25.

[24] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting,* Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[25] P. L. COMBETTES AND I. YAMADA, *Compositions and convex combinations of averaged nonexpansive operators,* J. Math. Anal. Appl., 425 (2015), pp. 55–70.

[26] D. DAVIS, *Convergence rate analysis of primal-dual splitting schemes,* SIAM J. Optim., 25 (2015), pp. 1912–1943.

[27] J. B. DIAZ AND F. T. METCALF, *On the structure of the set of subsequential limit points of successive approximations,* Bull. Amer. Math. Soc., 73 (1967), pp. 516–519.

[28] W. G. DOTSON, *On the Mann iterative process,* Trans. Amer. Math. Soc., 149 (1970), pp. 65–73.

[29] L. FEJÉR, *Untersuchungen über Fouriersche Reihen,* Math. Ann., 58 (1903), pp. 51–69.

[30] C. W. GROETSCH, *A note on segmenting Mann iterates,* J. Math. Anal. Appl., 40 (1972), pp. 369–372.

[31] K. KNOPP, *Theory and Application of Infinite Series,* Blackie & Son, London, 1954.

[32] P. KÜGLER AND A. LEITÃO, *Mean value iterations for nonlinear elliptic Cauchy problems,* Numer. Math., 96 (2003), pp. 269–293.

[33] D. A. LORENZ AND T. POCK, *An inertial forward-backward algorithm for monotone inclusions,* J. Math. Imaging Vision, 51 (2015), pp. 311–325.

[34] W. R. MANN, *Mean value methods in iteration,* Proc. Amer. Math. Soc., 4 (1953), pp. 506–510.

[35] W. R. MANN, *Averaging to improve convergence of iterative processes,* Lecture Notes in Math., 701 (1979), pp. 169–179.

[36] J. J. MOREAU, *Fonctions convexes duales et points proximaux dans un espace hilbertien,* C. R. Acad. Sci. Paris Sér. A, 255 (1962), pp. 2897–2899.

[37] C. OUTLAW AND C. W. GROETSCH, *Averaging iteration in a Banach space,* Bull. Amer. Math. Soc., 75 (1969), pp. 430–432.

[38] B. T. POLYAK, *Gradient methods for minimizing functionals,* USSR Comput. Math. Math. Phys., 3 (1963), pp. 864–878.

[39] B. T. POLYAK, *Some methods of speeding up the convergence of iteration methods,* USSR Comput. Math. Math. Phys., 4 (1964), pp. 1–17.

[40] B. T. POLYAK, *Minimization of unsmooth functionals,* USSR Comput. Math. Math. Phys., 9 (1969), pp. 14–29.

[41] B. T. POLYAK, *Comparison of the speed of convergence of single-step and multistep algorithms of optimization when there are noises,* Engrg. Cybernetics, 15 (1977), pp. 6–10.

[42] B. E. RHOADES, *Fixed point iterations using infinite matrices,* Trans. Amer. Math. Soc., 196 (1974), pp. 161–176.

[43] K. SLAVAKIS AND I. YAMADA, *The adaptive projected subgradient method constrained by families of quasi-nonexpansive mappings and its application to online learning,* SIAM J. Optim., 23 (2013), pp. 126–152.

[44] O. TOEPLITZ, *Über allgemeine lineare Mittelbildungen,* Prace Mat.-Fiz., 22 (1911), pp. 113–119.

Chapitre 3

Matrices concentrantes, superconcentrantes, et méthodes inertielles

Nous introduisons une notion plus fine de matrice concentrante (cf. Définition 1.9) pour obtenir des résultats de sommabilité dans les méthodes de point fixe avec mémoire. La construction que nous proposons permet en particulier des pondérations négatives. La combinaison de ces résultats et du formalisme introduit dans le Chapitre 2, ouvre la voie à la construction de nouveaux algorithmes avec mémoire. [1]

3.1 Introduction

La notion de matrice concentrante a été développée dans [5] pour étudier la convergence d'une famille de schémas itératifs de point fixe avec mémoire. Nous étendons ici la définition de matrice concentrante à des matrices n'étant pas nécessairement à valeurs positives. La bibliographie du Chapitre 2 contient une liste de références sur les algorithmes avec mémoire. Le lecteur est invité à s'y reporter.

Définition 3.1 Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels telle que

$$\sup_{n\in\mathbb{N}}\sum_{j=0}^{n}|\mu_{n,j}|<+\infty,\quad \lim_{n\to+\infty}\sum_{j=0}^{n}\mu_{n,j}=1\quad\text{et}\quad(\forall j\in\mathbb{N})\quad\lim_{n\to+\infty}\mu_{n,j}=0. \tag{3.1}$$

Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est *concentrante* s'il existe une suite $(\chi_n)_{n\in\mathbb{N}}$ à valeurs dans $]0,+\infty[$

1. Un article basé sur ce chapitre est en préparation.

telle que $\inf_{n\in\mathbb{N}} \chi_n > 0$ et si toute suite $(\xi_n)_{n\in\mathbb{N}}$ à valeurs dans $[0,+\infty[$ qui satisfait

$$\left(\exists\,(\varepsilon_n)_{n\in\mathbb{N}} \in [0,+\infty[^{\mathbb{N}}\right) \sum_{n\in\mathbb{N}} \chi_n\varepsilon_n < +\infty \quad \text{et} \quad (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j + \varepsilon_n \tag{3.2}$$

converge. De plus, $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est *superconcentrante* s'il existe une suite $(\chi_n)_{n\in\mathbb{N}}$ à valeurs dans $]0,+\infty[$ telle que $\inf_{n\in\mathbb{N}} \chi_n > 0$ et si toutes suites $(\xi_n)_{n\in\mathbb{N}}$ et $(\beta_n)_{n\in\mathbb{N}}$ à valeurs dans $[0,+\infty[$ qui satisfont

$$\left(\exists\,(\varepsilon_n)_{n\in\mathbb{N}} \in [0,+\infty[^{\mathbb{N}}\right) \sum_{n\in\mathbb{N}} \chi_n\varepsilon_n < +\infty \quad \text{et} \quad (\forall n \in \mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j - \beta_n + \varepsilon_n \tag{3.3}$$

respectivement converge et est dans $\ell^1(\mathbb{N})$.

Définition 3.2 Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels vérifiant (3.1) et soit $p \in \mathbb{N}\smallsetminus\{0\}$. L'ordre d'une matrice vaut p si pour tous les entiers $n \geqslant p$ et $j \leqslant n-p$, $\mu_{n,j} = 0$. Si aucun entier p ne satisfait cette propriété, la matrice est dite d'ordre infini.

3.2 Résultats préliminaires

Commençons par quelques lemmes utiles sur des suites réelles.

Lemme 3.3 *Soit $j \in \mathbb{N}$, soit $(\nu_n)_{n\in\mathbb{N}}$ à valeurs dans $[0,1[$, soit $(\theta_n)_{n\in\mathbb{N}}$ à valeurs dans $]0,+\infty[$ et soient $(\sigma_n)_{n\in\mathbb{N}}$ et $(\eta_n)_{n\in\mathbb{N}}$ deux suites à valeurs dans $\mathbb{R}_+$ telles que*

$$(\forall n \geqslant j) \quad \theta_{n+2} \leqslant \eta_{n+2}\theta_{n+1} + \sigma_{n+2}\theta_n. \tag{3.4}$$

Supposons que $(\forall n \geqslant j)$ $\eta_{n+2}\nu_{n+1} + \sigma_{n+2} \leqslant \nu_{n+1}\nu_{n+2}$ et que $\theta_j \leqslant \nu_j$. Alors $(\forall n \geqslant j)$ $\theta_n \leqslant \prod_{l=j}^{n} \nu_l$.

Démonstration. On procède par récurrence forte. Soit $n \geqslant j$, soit $k \in \{j,\ldots,n+1\}$ et on suppose que $\theta_k \leqslant \prod_{l=j}^{k} \nu_l$. Alors par (3.4), $\theta_{n+2} \leqslant \eta_{n+2}\prod_{l=j}^{n+1}\nu_l + \sigma_{n+2}\prod_{l=j}^{n}\nu_l$. On remarque que

$$\eta_{n+2}\nu_{n+1} + \sigma_{n+2} \leqslant \nu_{n+1}\nu_{n+2} \Rightarrow \eta_{n+2}\prod_{l=j}^{n+1}\nu_l + \sigma_{n+2}\prod_{l=j}^{n}\nu_l \leqslant \prod_{l=j}^{n+2}\nu_l \tag{3.5}$$

et ainsi,

$$\theta_{n+2} \leqslant \prod_{l=j}^{n+2}\nu_l, \tag{3.6}$$

qui conclut la récurrence. □

Lemme 3.4 [8, §43.3.] *Soient* $(\alpha_n)_{n\in\mathbb{N}}$ *et* $(\beta_n)_{n\in\mathbb{N}}$ *deux suites à valeurs dans* $\mathbb{R}$. *Supposons que* $\sum_{n\in\mathbb{N}} |\alpha_n - \alpha_{n+1}| < +\infty$ *et que* $\sum_{n\in\mathbb{N}} \beta_n < +\infty$. *Alors* $\sum_{n\in\mathbb{N}} \alpha_n\beta_n < +\infty$.

Lemme 3.5 [8, Theorem 43.5] *Soit* $(\xi_n)_{n\in\mathbb{N}}$ *une suite réelle et soit* $\xi \in \mathbb{R}$. *Supposons que* $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *est une matrice concentrante de réels. Alors* $\xi_n \to \xi \Rightarrow \sum_{j=0}^{n} \mu_{n,j}\xi_j \to \xi$.

Lemme 3.6 [3, Lemma 4.5] *Soit* $j \in \mathbb{N}$. *On a*

$$\sum_{k=j}^{+\infty}\prod_{l=j}^{k}\frac{l-1}{l+2} \leqslant \frac{j+5}{2} \tag{3.7}$$

Lemme 3.7 *Soient* $(\sigma_n)_{n\in\mathbb{N}}$ *et* $(\eta_n)_{n\in\mathbb{N}}$ *deux suites à valeurs dans* $]0,1]$. *Soit* $(\theta_n)_{n\in\mathbb{N}}$ *une suite à valeurs* $[0,+\infty[$ *telle que* $\theta_{-1} = \theta_0 = 0$ *et telle qu'il existe une suite* $(\varepsilon_n)_{n\in\mathbb{N}}$ *dans* $[0,+\infty[$ *qui satisfait*

$$(\forall n \in \mathbb{N}) \quad \theta_{n+1} \leqslant \eta_n\theta_n + \sigma_n\theta_{n-1} + \varepsilon_n. \tag{3.8}$$

On pose

$$u_1 = \begin{bmatrix}1\\0\end{bmatrix}, \quad (\forall n \in \mathbb{N}) \quad A_n = \begin{bmatrix}\eta_n & \sigma_n\\ 1 & 0\end{bmatrix}, e_n = \begin{bmatrix}\varepsilon_n\\0\end{bmatrix} \tag{3.9}$$

et

$$(\forall l \in \mathbb{N})(\forall n \geqslant l) \quad \mathcal{A}_l^n = A_n\cdots A_l, \quad \tau_{n,l} = u_1^\top \mathcal{A}_l^n u_1 \quad \text{et} \quad \chi_n = \sum_{l\geqslant n}\tau_{n,l+1}. \tag{3.10}$$

Supposons que $\sum_{j\geqslant 0}\varepsilon_j\chi_j < +\infty$. *Alors* $\sum_{n\geqslant 0}\theta_{n+1} \leqslant \sum_{j\geqslant 0}\varepsilon_j\chi_j < +\infty$.

Démonstration. On définit

$$(\forall n \in \mathbb{N}) \quad t_n = \begin{bmatrix}\theta_{n+1}\\ \theta_n\end{bmatrix} \quad \text{et} \quad e_n = \begin{bmatrix}\varepsilon_n\\0\end{bmatrix}. \tag{3.11}$$

Ainsi, en utilisant $\preccurlyeq$ pour l'ordre composante par composante, on obtient

$$(\forall n \in \mathbb{N}) \quad t_{n+1} \preccurlyeq A_n t_n + e_n \tag{3.12}$$

et par induction $(\forall n \in \mathbb{N})$ $t_{n+1} \preccurlyeq \sum_{j=0}^{n}\mathcal{A}_{j+1}^n e_j$. On a $(\forall n \in \mathbb{N})$ $\theta_{n+1} \leqslant u_1^\top\sum_{j=0}^{n}\mathcal{A}_{j+1}^n e_j = \sum_{j=0}^{n}u_1^\top\mathcal{A}_{j+1}^n e_j$ et $(\forall j \in \{0,\dots,n\})$ $u_1^\top\mathcal{A}_{j+1}^n e_j = \varepsilon_j u_1^\top\mathcal{A}_{j+1}^n u_1$. Ainsi,

$$\sum_{n\geqslant 0}\theta_{n+1} \leqslant \sum_{n\geqslant 0}\sum_{j=0}^{n}u_1^\top\mathcal{A}_{j+1}^n e_j = \sum_{j\geqslant 0}\varepsilon_j\sum_{n\geqslant j}u_1^\top\tau_{n,j+1}u_1 = \sum_{j\geqslant 0}\varepsilon_j\chi_j, \tag{3.13}$$

ce qui conclut la démonstration. □

Remarque 3.8 Supposons qu'il existe η et σ dans $[0,1[$ tels que $\eta+\sigma<1$ et $(\forall n\in\mathbb{N})$ $\sigma_n<\sigma$, $\eta_n<\eta$. On pose $\nu=(\eta+\sqrt{\eta^2+4\sigma})/2<1$. En remarquant que $(\forall j\in\mathbb{N})(\forall n\geqslant j)$ $\tau_{n+2,j}\leqslant\eta\tau_{n+1,j}+\sigma\tau_{n,j}$, et que $\eta\nu+\sigma\leqslant\nu^2$, on peut appliquer le Lemme 3.3 et conclure facilement que $(\chi_n)_{n\in\mathbb{N}}$ est bornée.

Lemme 3.9 *Soit $(\xi_n)_{n\in\mathbb{N}}$ une suite de réels positifs. On pose $\theta_{-1}=\theta_0=0$ et $(\forall n\in\mathbb{N})$ $\theta_{n+1}=[\xi_{n+1}-\xi_n]^+$. Supposons que $\sum_{n\geqslant 0}\theta_{n+1}<+\infty$. Alors $(\xi_n)_{n\in\mathbb{N}}$ converge.*

Démonstration. On pose $(\forall n\in\mathbb{N})$ $\omega_n=\xi_n-\sum_{j=0}^n\theta_j$. On sait que $\inf_{n\in\mathbb{N}}\xi_n\geqslant 0$, donc $(\omega_n)_{n\in\mathbb{N}}$ est bornée inférieurement et

$$(\forall n\in\mathbb{N})\quad \omega_{n+1}=\xi_{n+1}-\theta_{n+1}-\sum_{j=0}^n\theta_j\leqslant\xi_{n+1}-\xi_{n+1}+\xi_n-\sum_{j=0}^n\theta_j=\omega_n. \tag{3.14}$$

Ainsi $(\omega_n)_{n\in\mathbb{N}}$ converge, et donc $(\xi_n)_{n\in\mathbb{N}}$ aussi. □

Les deux lemmes suivants serviront à étudier la convergence forte de nos algorithmes, mais aussi des cas particuliers de techniques inertielles.

Lemme 3.10 *Soient β et γ dans $]0,1[$ tels que $\beta+2\gamma<1$. Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels d'ordre 3 telle que*

$$(\forall n\in\{2,3,\ldots\})\quad \mu_{n,n-1}<\beta\quad \text{et}\quad \mu_{n,n-2}<\gamma. \tag{3.15}$$

Soit $(x_n)_{n\in\mathbb{N}}$ une suite de $\mathcal{H}$ telle que $\sum_{n\in\mathbb{N}}\|x_{n+1}-\sum_{j=0}^n\mu_{n,j}x_j\|<+\infty$. Alors $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy.

Démonstration. On veut montrer que $\sum_{n\in\mathbb{N}}\|x_{n+1}-x_n\|<+\infty$. On pose $(\forall n\in\mathbb{N})$ $\overline{x}_n=\sum_{j=0}^n\mu_{n,j}x_j$. On a

$$\begin{aligned}(\forall n\geqslant 2)\quad \|x_{n+1}-x_n\| &\leqslant \|x_{n+1}-\overline{x}_n\|+\|\overline{x}_n-x_n\|\\ &\leqslant \|(\mu_{n,n-1}+\mu_{n,n-2})(x_{n-1}-x_n)+\mu_{n,n-2}(x_{n-2}-x_{n-1})\|\\ &\quad +\|x_{n+1}-\overline{x}_n\|.\end{aligned} \tag{3.16}$$

On pose

$$(\forall n\in\mathbb{N})\quad \theta_{n+1}=\|x_{n+1}-x_n\|\quad\text{et}\quad \varepsilon_n=\|x_{n+1}-\overline{x}_n\|. \tag{3.17}$$

Il suit que

$$\begin{aligned}(\forall n\in\{2,3,\ldots\})\quad \theta_{n+1} &\leqslant (\mu_{n,n-1}+\mu_{n,n-2})\theta_n+\mu_{n,n-2}\theta_{n-1}+\varepsilon_n\\ &\leqslant (\beta+\gamma)\theta_n+\gamma\theta_{n-1}+\varepsilon_n.\end{aligned} \tag{3.18}$$

On utilise les notations du Lemme 3.7. On a $(\chi_n)_{n\in\mathbb{N}}$ bornée d'après la Remarque 3.8 et on appelle $\chi\in]0,+\infty[$ sa borne. Ainsi le Lemme 3.7 assure que

$$\sum_{n\geqslant 0}\theta_{n+1}\leqslant\sum_{j\geqslant 0}\varepsilon_j\chi_j\leqslant\chi\sum_{j\geqslant 0}\varepsilon_j<+\infty. \tag{3.19}$$

On conclut ainsi que $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy. □

Proposition 3.11 *Soient β et γ dans $]0,1[$ tels que $\beta + 2\gamma < 1$. Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice d'ordre 3 superconcentrante telle que*

$$(\forall n \in \mathbb{N}) \quad \mu_{n,n-1} < \beta, \quad \mu_{n,n-2} < \gamma, \tag{3.20}$$

et soient $(x_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathcal{H}$ et S un sous-ensemble d'intérieur non vide de $\mathcal{H}$ tels que

$$(\forall x \in S)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leqslant \|\overline{x}_n - x\| + \varepsilon_n. \tag{3.21}$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge fortement.

Démonstration. Soient $x \in \operatorname{int} S$ et $\rho \in]0,+\infty[$ tels que $B(x,\rho) \subset S$. On définit $(y_n)_{n\in\mathbb{N}}$ à valeurs dans $B(x,\rho)$ par

$$(\forall n \in \mathbb{N}) \quad y_n = \begin{cases} x, & \text{si} \quad x_{n+1} = \overline{x}_n; \\ x - \rho\dfrac{x_{n+1} - \overline{x}_n}{\|x_{n+1} - \overline{x}_n\|}, & \text{sinon.} \end{cases}$$

Ainsi, on a $(\forall n \in \mathbb{N})\ \|x_{n+1} - y_n\| \leqslant \|\overline{x}_n - y_n\| + \varepsilon_n$. De plus

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leqslant \|\overline{x}_n - x\| - 2\rho\|x_{n+1} - \overline{x}_n\| + \varepsilon_n. \tag{3.22}$$

et $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ étant superconcentrante, on a $\sum_{n\in\mathbb{N}} \|x_{n+1} - \overline{x}_n\| < +\infty$. Donc par le Lemme 3.10, $(x_n)_{n\in\mathbb{N}}$ est une suite de Cauchy, et donc converge fortement. □

3.3 Construction de matrices superconcentrantes

3.3.1 Matrices superconcentrantes positives

Le résultat suivant donne un critère pour s'assurer qu'une matrice concentrante positive d'ordre fini est superconcentrante.

Théorème 3.12 *Soit p un entier strictement positif et soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice concentrante d'ordre p de réels positifs. Supposons que pour tout entier $j < p$, $\sum_{n\leqslant p} |\mu_{n+1,n+1-j} - \mu_{n,n-j}| < +\infty$. Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante.*

Démonstration. On définit

$$(\forall n \in \mathbb{N})(\forall k \in \{0,\ldots,n\}) \quad \tau_{n,k} = \sum_{j=0}^{k} \mu_{n,j}. \tag{3.23}$$

Supposons que $(\xi_n)_{n\in\mathbb{N}}$ est une suite à valeurs dans $[0,+\infty[$ telle qu'il existe deux suites $(\beta_n)_{n\in\mathbb{N}}$ et $(\varepsilon_n)_{n\in\mathbb{N}}$ dans $[0,+\infty[$ qui satisfont

$$\sum_{n\in\mathbb{N}} \varepsilon_n < +\infty \quad \text{et} \quad (\forall n\in\mathbb{N}) \quad \xi_{n+1} \leqslant \sum_{j=0}^{n} \mu_{n,j}\xi_j - \beta_n + \varepsilon_n. \tag{3.24}$$

On sait que $(\xi_n)_{n\in\mathbb{N}}$ converge. Pour montrer que $\sum_{n\in\mathbb{N}} \beta_n < +\infty$, il suffit de prouver que $\sum_{n\geqslant p} \Big(\sum_{j=0}^{n} \mu_{n,j}\xi_j - \xi_{n+1} + \varepsilon_n\Big) < +\infty$. De plus,

$$\begin{aligned}(\forall n\geqslant p) \quad \sum_{j=0}^{n} \mu_{n,j}\xi_j - \xi_{n+1} + \varepsilon_n &= \sum_{j=0}^{n} \mu_{n,j}\Bigg(\sum_{k=j}^{n}(\xi_k - \xi_{k+1})\Bigg) + \varepsilon_n \\ &= \sum_{k=0}^{n}\Bigg(\sum_{j=0}^{k}\mu_{n,j}\Bigg)(\xi_k - \xi_{k+1}) + \varepsilon_n \\ &= \sum_{k=0}^{n} \tau_{n,k}(\xi_k - \xi_{k+1}) + \varepsilon_n. \end{aligned} \tag{3.25}$$

On remarque que pour tous entiers $n \geqslant p$ et $k \leqslant n-p$, on a $\tau_{n,k} = 0$. Ainsi, pour tout entier $n \geqslant p$,

$$\begin{aligned}\sum_{k=0}^{n} \tau_{n,k}(\xi_k - \xi_{k+1}) &= \sum_{k=n-p+1}^{n} \tau_{n,k}(\xi_k - \xi_{k+1}) \\ &= \sum_{k=1}^{p} \tau_{n,k+n-p}(\xi_{k+n-p} - \xi_{k+n-p+1}). \end{aligned} \tag{3.26}$$

Pour tout entier strictement positif $k \leqslant p$, on étudie indépendamment les p suites $(\tau_{n,k+n-p}(\xi_{k+n-p} - \xi_{k+n-p+1}))_{n\in\mathbb{N}}$. Par le Lemme 3.4, il reste à montrer que pour tout entier strictement positif $k \leqslant p$, on a $\sum_{n\geqslant p} |\tau_{n+1,k+n-p+1} - \tau_{n,k+n-p}| < +\infty$. Or, pour tous entiers $n \geqslant p$ et $k \leqslant p$ strictement positifs,

$$
\begin{aligned}
\left|\tau_{n+1,k+n-p+1} - \tau_{n,k+n-p}\right| &= \left|\sum_{j=0}^{k+n-p+1} \mu_{n+1,j} - \sum_{j=0}^{k+n-p} \mu_{n,j}\right| \\
&= \left|\sum_{j=n-p+2}^{k+n-p+1} \mu_{n+1,j} - \sum_{j=n-p+1}^{k+n-p} \mu_{n,j}\right| \\
&= \left|\sum_{j=n-p+2}^{k+n-p+1} \mu_{n+1,j} - \mu_{n,j-1}\right| \\
&\leqslant \sum_{j=n-p+2}^{k+n-p+1} \left|\mu_{n+1,j} - \mu_{n,j-1}\right| \\
&\leqslant \sum_{j=n-p+2}^{n+1} \left|\mu_{n+1,j} - \mu_{n,j-1}\right| \\
&= \sum_{j=0}^{p-1} \left|\mu_{n+1,n+1-j} - \mu_{n,n-j}\right|, \qquad (3.27)
\end{aligned}
$$

ce qui conclut la démonstration grâce à l'hypothèse de départ. □

La conjecture suivante est une extension du théorème précédent pour des matrices d'ordre infini.

Conjecture 3.13 *Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice concentrante de réels positifs. On pose*

$$
(\forall n \in \mathbb{N})(\forall j \in \mathbb{N}) \quad \tau_{n,j} = \begin{cases} \left|\mu_{n+1,n+1-j} - \mu_{n,n-j}\right|, & \textit{si } n \geqslant j; \\ 0, & \textit{si } n < j. \end{cases} \qquad (3.28)
$$

Supposons que pour tout entier $j \in \mathbb{N}$, $\sum_{n\in\mathbb{N}} \tau_{n,j} < +\infty$. Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante.

Exemple 3.14 Soit $p \in \mathbb{N} \setminus \{0\}$ et soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice d'ordre $p+1$ de réels positifs. Soit $(a_i)_{0\leqslant i\leqslant p}$ une famille à valeurs dans $[0,1]$ telle que $\sum_{i=0}^{p} a_i = 1$ et telle que les racines du polynôme $z \mapsto z^{p+1} - \sum_{j=0}^{p} a_j z^{p-j}$ sont toutes dans le disque unité fermé avec exactement une racine sur sa frontière. Supposons que

$$
\begin{cases}
(\forall n \in \{0,\ldots,p-1\})(\forall j \in \{0,\ldots,n\}) \quad \mu_{n,j} = \begin{cases} 0 & \text{if} \quad 0 \leqslant j \leqslant n, \\ 1 & \text{if} \quad j = n, \end{cases} \\
(\forall n \geqslant p)(\forall j \in \{0,\ldots,n\}) \quad \mu_{n,j} = \begin{cases} 0 & \text{if} \quad 0 \leqslant j < n-p, \\ a_{n-j} & \text{if} \quad n-p \leqslant j \leqslant n. \end{cases}
\end{cases} \qquad (3.29)
$$

Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante.

Démonstration. $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est concentrante par [5, Example 2.7]. On remarque que $(\forall n \geqslant p)(\forall j \in \{0,\ldots,n\})\ \mu_{n,n-j} = a_j$. Il suffit alors d'appliquer le Théorème 3.12 pour conclure. □

Exemple 3.15 Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels positifs d'ordre 2 et soit $(\alpha_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,1]$ telle que $\underline{\lim}\,\alpha_n > 0$. Supposons que $(\forall n \in \mathbb{N})\ \mu_{n,n} = \alpha_n$ et $\mu_{n,n-1} = 1-\alpha_n$. Supposons que $((1-\alpha_n)(1-\alpha_{n+1}))_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$ et que $(|\alpha_{n+1}-\alpha_n|) \in \ell^1(\mathbb{N})$. Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante.

Démonstration. $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est concentrante par [5, Example 2.5] et car $((1-\alpha_n)(1-\alpha_{n+1}))_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$. Il suffit alors de remarquer que $(\forall n \in \mathbb{N})\ \mu_{n,n} - \mu_{n-1,n-1} = \alpha_{n+1} - \alpha_n$ et d'appliquer le Théorème 3.12. □

Remarque 3.16 Les deux conditions ne sont pas redondantes. On prend $\alpha_n = 1-1/\sqrt{n}$. Dans ce cas $((1-\alpha_n)(1-\alpha_{n+1}))_{n\in\mathbb{N}} \notin \ell^1(\mathbb{N})$ et $(|\alpha_{n+1}-\alpha_n|)_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$. On prend maintenant $\alpha_n = 1-((-1)^n/n+2/n)$. Dans ce cas, on obtient le résultat inverse.

Remarque 3.17 Supposons que $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est d'ordre 1. On retrouve alors le cas sans mémoire et la matrice est bien superconcentrante d'après l'Exemple 3.15.

Remarque 3.18 Soient N et p dans $\mathbb{N}\smallsetminus\{0\}$, soit $T\colon \mathbb{R}^N \to \mathbb{R}^N$ une contraction, et soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice d'ordre p qui vérifie (3.1) et telle que ses coefficients sont choisis de la manière suivante

$$\begin{array}{l}\text{pour } n = 0,1,\ldots\\ \left\lfloor\begin{array}{l}\underset{(\mu_{n,j})_{1\leqslant j\leqslant p}\in[0,1]^p}{\text{Argmin}}\ \left\|\displaystyle\sum_{j=n-p+1}^{n}\mu_{n,j}(Tx_j-x_j)\right\|_2^2\\ x_{n+1} = \displaystyle\sum_{j=0}^{n}\mu_{n,j}Tx_j.\end{array}\right.\end{array} \tag{3.30}$$

Cette méthode de sélection est connue sous le nom de *méthode d'accélération d'Anderson* [9]. Il est naturel de se demander si les poids produits par l'accélération d'Anderson engendrent une matrice concentrante, voire superconcentrante. Dans ce cas nous pourrions appliquer cette méthode à nos problèmes. Tel que démontré ultérieurement dans la Section 3.4, notre formalisme est également capable de traiter le cas d'une combinaison affine des images plutôt que des itérés.

3.3.2 Matrices inertielles d'ordre 3

Par souci de clarté, nous avons choisi de présenter les résultats de cette section pour des matrices d'ordre 3. Ces résultats s'étendent à des matrices d'ordre $p \geqslant 2$ quelconque.

Théorème 3.19 *Soient* $(\sigma_n)_{n\in\mathbb{N}}$ *et* $(\eta_n)_{n\in\mathbb{N}}$ *deux suites à valeurs dans* $]0,1]$ *et soit* $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *une matrice de réels d'ordre* 3 *telle que* $\mu_{0,0}=1$*,* $\mu_{1,0}=0$*,* $\mu_{1,1}=1$ *et*

$$(\forall n\in\{2,3,\ldots\})(\forall j\in\mathbb{N})\quad \mu_{n,j}=\begin{cases}1+\eta_n, & \textit{si } j=n;\\ \sigma_n-\eta_n, & \textit{si } j=n-1;\\ -\sigma_n, & \textit{si } j=n-2.\end{cases}\tag{3.31}$$

Supposons que

$$(\forall j\in\mathbb{N})\quad \eta_j\leqslant\nu_j\quad \textit{et}\quad (\forall n\in\mathbb{N})\quad \eta_{n+1}\nu_{n+1}+\sigma_{n+1}\leqslant\nu_{n+1}\nu_{n+2}\tag{3.32}$$

et que l'une des propriétés suivantes est satisfaite :

(i) $\sum_{n\geqslant j}\prod_{l=j}^{n}\nu_l<+\infty$.

(ii) *Il existe* $a\in\,]2,+\infty[$ *tel que* $(\forall n\in\mathbb{N})$ $\nu_n=(n-1)/(n+a)$.

(iii) *Il existe* $\nu\in\,]0,1[$ *tel que* $(\forall n\in\mathbb{N})$ $\nu_n<\nu$.

Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ *est superconcentrante.*

Démonstration. On utilise les notations du Lemme 3.7, et on traite en premier lieu la concentration.

(i) : D'abord remarquons que $(\forall j\in\mathbb{N})(\forall n\geqslant j)$ $\tau_{n+2,j}=\eta_{n+2}\tau_{n+2,j}+\sigma_{n+2}\tau_{n,j}$. Ainsi, par le Lemme 3.3, $\tau_{n,j}\leqslant\prod_{l=j}^{n}\nu_l$, et donc $\chi_n<+\infty$. Maintenant, on suppose que $(\xi_n)_{n\in\mathbb{N}}$ est une suite de $[0,+\infty[$ telle qu'il existe une suite $(\varepsilon_n)_{n\in\mathbb{N}}$ dans $[0,+\infty[$ qui vérifie

$$\sum_{n\in\mathbb{N}}\chi_n\varepsilon_n<+\infty\quad\text{et}\quad(\forall n\in\mathbb{N})\quad \xi_{n+1}\leqslant\sum_{j=0}^{n}\mu_{n,j}\xi_j+\varepsilon_n.\tag{3.33}$$

Il reste à montrer que $(\xi_n)_{n\in\mathbb{N}}$ converge. Nous avons $\xi_1-\xi_0\leqslant(\mu_{0,0}-1)\xi_0+\varepsilon_0$ et

$$(\forall n\geqslant 2\quad \xi_{n+1}-\xi_n=\eta_n(\xi_n-\xi_{n-1})+\sigma_n(\xi_{n-1}-\xi_{n-2})+\varepsilon_n.\tag{3.34}$$

On pose alors $\theta_{-1}=\theta_0=0$ et $(\forall n\in\mathbb{N})$ $\theta_{n+1}=[\xi_{n+1}-\xi_n]^+$. On déduit de (3.33) et des inégalités ci-dessus que $(\forall n\in\mathbb{N})$ $\theta_{n+1}\leqslant\eta_n\theta_n+\sigma_n\theta_{n-1}+\varepsilon_n$. Il suit du Lemme 3.7 que $\sum_{n\geqslant0}\theta_{n+1}\leqslant\sum_{j\geqslant0}\varepsilon_j\chi_j$. On sait par (3.33) que $\sum_{j\in\mathbb{N}}\chi_j\varepsilon_j<+\infty$, donc $\sum_{n\in\mathbb{N}}\theta_n<+\infty$. Finalement, la convergence de $(\xi_n)_{n\in\mathbb{N}}$ est donnée par le Lemme 3.9.

(ii)&(iii) : (iii)⇒(i) est claire et il suffit de remarquer que (ii)⇒(i) par le Lemme 3.6 pour conclure. Pour la superconcentration, on appelle $(\beta_n)_{n\in\mathbb{N}}$ la suite qui doit être sommable. Elle vérifie

$$(\forall n\in\mathbb{N})\quad \theta_{n+2}\leqslant\eta_n\theta_{n+1}+\sigma_n\theta_n-\beta_n+\varepsilon_n.\tag{3.35}$$

La positivité de $(\theta_n)_{n\in\mathbb{N}}$ et la sommabilité de $(\varepsilon_n)_{n\in\mathbb{N}}$, $(\eta_n\theta_{n+1})_{n\in\mathbb{N}}$ et $(\sigma_n\theta_n)_{n\in\mathbb{N}}$ permet de conclure. □

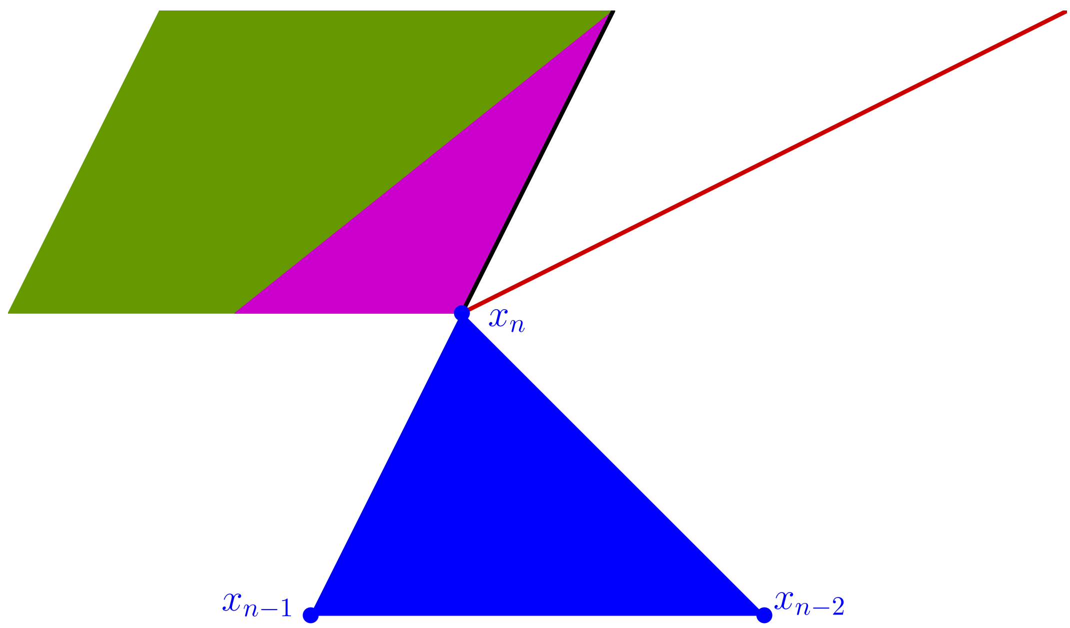


FIGURE 3.1 – Zones de concentration

Comme illustré par la Figure 3.1, les matrices concentrantes permettent d'agir sur l'enveloppe affine de l'orbite des itérés. Le Théorème 3.19 assure que l'on peut agir dans la zone verte. La zone violette correspond au cas où $(\forall n \in \mathbb{N})$ $\eta_n < \eta$, $\sigma_n < \sigma$ et $\eta + \sigma < 1$ dans la Remarque 3.8. Les moyennes de Mann permettent, elles, de traiter la zone bleue. L'inertiel classique correspond simplement à la ligne jaune. Les cas compliqués traités par le théorème sont les frontières des différentes zones. La conjecture suivante offrirait la possibilité d'action sur le segment rouge. Une telle possibilité démontrerait la flexibilité de cette théorie.

Conjecture 3.20 *Soit $(\eta_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,1[$, et soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice de réels telle que $\mu_{0,0} = 1$, $\mu_{1,0} = 0$, $\mu_{1,1} = 0$ et*

$$(\forall n \in \{2,3,\ldots\})(\forall j \in \mathbb{N}) \quad \mu_{n,j} = \begin{cases} 1+\eta_n, & \textit{si } \ j = n; \\ -2\eta_n, & \textit{si } \ j = n-1; \\ \eta_n, & \textit{si } \ j = n-2; \\ 0, & \textit{si } \ j < n-2. \end{cases} \tag{3.36}$$

Supposons que $(\forall n \in \mathbb{N})$ $\eta_n \leqslant (n-1)/(n+3)$. Alors $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante.

3.4 Applications

Nous considérons le même problème qu'au Chapitre 2.

Problème 3.21 Soit m un entier strictement positif. Pour tout $n \in \mathbb{N}$ et tout $i \in \{1, \ldots, m\}$, soit $\alpha_{i,n} \in \left]0,1\right]$ et soit $T_{i,n} \colon \mathcal{H} \to \mathcal{H}$ un opérateur $\alpha_{i,n}$-moyenné si $i < m$, et $\alpha_{m,n}$-quasimoyenné si $i = m$. De plus, pour tout $n \in \mathbb{N}$, l'une des propriétés suivantes doit être satisfaite :

(a) $T_{m,n}$ est $\alpha_{m,n}$-moyenné.

(b) $m > 1$, $\alpha_{m,n} < 1$ et $\bigcap_{i=1}^{m} \operatorname{Fix} T_{i,n} \neq \varnothing$.

(c) $m = 1$.

L'objectif est de

$$\text{trouver} \quad x \in S = \bigcap_{n \in \mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad \text{où} \quad (\forall n \in \mathbb{N}) \quad T_n = T_{1,n} \cdots T_{m,n}. \tag{3.37}$$

On va utiliser le formalisme habituel de la thèse pour le résoudre, mais avec des matrices superconcentrantes.

Algorithme 3.22 On considère le cadre du Problème 3.21. Pour tout $n \in \mathbb{N}$, soit ϕ_n une constante de moyennage de T_n, soit $\lambda_n \in \left]0, 1/\phi_n\right]$ et pour tout $i \in \{1, \ldots, m\}$, soit $e_{i,n} \in \mathcal{H}$. Soit $(\mu_{n,j})_{n \in \mathbb{N}, 0 \leqslant j \leqslant n}$ une matrice superconcentrante. Soit $x_0 \in \mathcal{H}$. On itère

$$\begin{array}{l} \text{pour } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\ x_{n+1} = \overline{x}_n + \lambda_n \Big(T_{1,n} \Big(T_{2,n} \big(\cdots T_{m-1,n} (T_{m,n} \overline{x}_n + e_{m,n}) + e_{m-1,n} \cdots \big) + e_{2,n} \Big) + e_{1,n} - \overline{x}_n \Big). \end{array} \right. \end{array} \tag{3.38}$$

Une extension du Théorème 2.3 du Chapitre 2 au cas superconcentrant est ici présentée. Désormais, ce théorème recouvre entièrement les résultats récents d'itérations d'opérateurs moyennés [6].

Théorème 3.23 *On considère le cadre de l'Algorithme 3.22. Pour tout $n \in \mathbb{N}$, on définit*

$$(\forall i \in \{1, \ldots, m\})(\forall n \in \mathbb{N}) \quad T_{i+,n} = \begin{cases} T_{i+1,n} \cdots T_{m,n}, & \textit{if } i \neq m; \\ \mathrm{Id}\,, & \textit{if } i = m \end{cases} \tag{3.39}$$

et on pose

$$\nu_n \colon S \to \left[0, +\infty\right[: x \mapsto \vartheta_n \big(2\|\overline{x}_n - x\| + \vartheta_n \big). \tag{3.40}$$

Supposons que

$$\sum_{n\in\mathbb{N}}\chi_n\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^-\|x_j-x_k\|^2<+\infty \quad et \quad (\forall x\in S)\quad \sum_{n\in\mathbb{N}}\chi_n\nu_n(x)<+\infty. \tag{3.41}$$

Alors :

(i) $\sum_{n\in\mathbb{N}}\sum_{j=0}^{n}\sum_{k=0}^{n}[\mu_{n,j}\mu_{n,k}]^+\|x_j-x_k\|^2<+\infty$.

(ii) $\sum_{n\in\mathbb{N}}\lambda_n(1/\phi_n-\lambda_n)\|T_{1,n}\cdots T_{m,n}\overline{x}_n-\overline{x}_n\|^2<+\infty$.

(iii) *Soit $x\in S$. Supposons que $(\exists\,\varepsilon\in\,]0,1[)(\forall n\in\mathbb{N})\ \lambda_n\leqslant\varepsilon+(1-\varepsilon)/\phi_n$. Alors*

$$\sum_{n\in\mathbb{N}}\lambda_n\max_{1\leqslant i\leqslant m}\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id}-T_{i,n})T_{i+,n}\overline{x}_n-(\mathrm{Id}-T_{i,n})T_{i+,n}x\|^2<+\infty.$$

(iv) $(x_n)_{n\in\mathbb{N}}$ *converge faiblement vers un point de S si tout point d'accumulation séquentiel faible de $(\overline{x}_n)_{n\in\mathbb{N}}$ est dans S. Dans ce cas, la convergence est forte si $\operatorname{int} S\neq\varnothing$ et si (3.20) est satisfaite.*

Démonstration. C'est un corollaire du Théorème 2.3 avec une hypothèse supplémentaire sur la matrice $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$.

(ii) : Par le Théorème 2.3(ii), on a

$$\|x_{n+1}-x\|^2\leqslant\sum_{j=0}^{n}\mu_{n,j}\|x_j-x\|^2-\lambda_n(1/\phi_n-\lambda_n)\|T_{1,n}\cdots T_{m,n}\overline{x}_n-\overline{x}_n\|^2+\nu_n(x). \tag{3.42}$$

Étant donné que $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante, le résultat est prouvé.

(iii) : On utilise le Théorème 2.3(iii) et le même schéma de démonstration que dans le Théorème 2.15(ix).

(iv) : On applique le Théorème 2.3(iv) pour la convergence faible. Le résultat de convergence forte est fourni par la Proposition 3.11 et par (3.42) car pour tout $x\in S$, $(\nu_n(x))_{n\in\mathbb{N}}$ est bornée indépendamment de x. ☐

Remarque 3.24 On considère le cadre du Théorème 3.23, en définissant cette fois-ci $(x_n)_{n\in\mathbb{N}}$ par

pour $n=0,1,\ldots$

$$\left\lfloor\begin{array}{l} y_n=x_n+\lambda_n\Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}x_n+e_{m,n})+e_{m-1,n}\cdots\big)+e_{2,n}\Big)+e_{1,n}-x_n\Big)\\ \overline{y}_n=\displaystyle\sum_{j=0}^{n}\mu_{n,j}y_j\\ x_{n+1}=\overline{y}_n. \end{array}\right.$$

$$(3.43)$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de S si tout point d'accumulation séquentiel faible de $(\overline{x}_n)_{n\in\mathbb{N}}$ est dans S.

Démonstration. On s'inspire de [5, Corollary 3.6] pour notre démonstration. On remarque que pour tout entier $n \geqslant 1$,

$$y_n = \overline{y}_{n-1} + \lambda_n\Big(T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}\overline{y}_{n-1}+e_{m,n})+e_{m-1,n}\cdots\big)+e_{2,n}\Big)+e_{1,n}-\overline{y}_{n-1}\Big). \tag{3.44}$$

Le Théorème 3.23(iv) assure que $(y_n)_{n\in\mathbb{N}}$ converge faiblement, on appelle $x \in S$ sa limite et on a $(\forall z \in \mathcal{H})$ $\langle y_n \mid z\rangle \to \langle x \mid z\rangle$. De plus, le Lemme 3.5 assure que $(\forall z \in \mathcal{H})$ $\langle \overline{y}_n \mid z\rangle \to \langle x \mid z\rangle$ et donc $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers x. □

Remarque 3.25 On considère le cadre du Théorème 3.23. Supposons que $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est d'ordre 1 et que

$$\lambda_n \in \left]0,(1-\varepsilon)(1+\varepsilon\phi_n)/\phi_n\right]. \tag{3.45}$$

Ce cas du Théorème 3.23 coïncide avec [6, Theorem 3.5].

Le théorème suivant est une version à deux pas de mémoire de l'algorithme de Krasnosel'skiĭ-Mann [7, 1]. L'objectif ici est double : d'une part il vise à exploiter le théorème précédent et ses résultats de sommabilité, et d'autre part il entend utiliser la propriété de concentration sur une autre suite que $(\|x_n - x\|)_{n\in\mathbb{N}}$.

Théorème 3.26 *Soit $T\colon \mathcal{H}\to\mathcal{H}$ une contraction telle que $\operatorname{Fix} T \neq \varnothing$, soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $[0,1]$ telle que $\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)=+\infty$, soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice superconcentrante de réels positifs d'ordre 2 et soit $x_0 \in \mathcal{H}$. On itère*

$$\begin{array}{l}\text{pour } n = 0,1,\ldots\\ \left\lfloor\begin{array}{l}\overline{x}_n = \displaystyle\sum_{j=0}^{n}\mu_{n,j}x_j\\ x_{n+1} = \overline{x}_n + \lambda_n\big(T\overline{x}_n - \overline{x}_n\big).\end{array}\right.\end{array} \tag{3.46}$$

Alors $\sum_{n\in\mathbb{N}}\lambda_n(1-\lambda_n)\|T\overline{x}_n-\overline{x}_n\|^2 < +\infty$. Supposons de plus que l'une des propriétés suivantes est satisfaite :

(a) *$(\|T\overline{x}_n - \overline{x}_n\|)_{n\in\mathbb{N}}$ converge.*

(b) *Il existe une suite $(\alpha_n)_{n\in\mathbb{N}}$ à valeurs dans $]0,1]$ telle que $\underline{\lim}\,\alpha_n > 0$ et $\sum_{n\geqslant 1}(1-\alpha_{n+1})(1-\alpha_n)(1-\alpha_{n-1}) < +\infty$ et $(\forall n\in\mathbb{N})$ $\mu_{n,n}=\alpha_n$ et $\mu_{n,n-1}=1-\alpha_n$.*

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Démonstration. Par le Théorème 3.23(ii), on sait que $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n)\|T\overline{x}_n-\overline{x}_n\|^2 < +\infty$. De plus, on sait que $\sum_{n\in\mathbb{N}} \lambda_n(1-\lambda_n) = +\infty$, donc on a $\underline{\lim}\,\|T\overline{x}_n - \overline{x}_n\| = 0$.

(a) : On en déduit que $\|T\overline{x}_n - \overline{x}_n\| \to 0$, et on conclut par demi-fermeture de $\operatorname{Id} - T$ [1, Theorem 4.27].

(b) : On pose

$$(\forall n \in \mathbb{N}) \quad \xi_n = \|T\overline{x}_n - \overline{x}_n\| \quad \text{et} \quad \delta_{n+1} = \|x_{n+1} - x_n\|. \tag{3.47}$$

Pour tout $n \in \mathbb{N}$,

$$\begin{aligned}
\|\overline{x}_{n+1} - T\overline{x}_{n+1}\| &\leqslant \|\overline{x}_{n+1} - T\overline{x}_n\| + \|T\overline{x}_n - T\overline{x}_{n+1}\| \\
&\leqslant \alpha_{n+1}\|x_{n+1} - T\overline{x}_n\| + (1-\alpha_{n+1})\|x_n - T\overline{x}_n\| \\
&\quad + \alpha_{n+1}\|x_{n+1} - \overline{x}_n\| + (1-\alpha_{n+1})\|x_n - \overline{x}_n\| \\
&= \alpha_{n+1}(1-\lambda_n)\|\overline{x}_n - T\overline{x}_n\| + \alpha_{n+1}\lambda_n\|T\overline{x}_n - \overline{x}_n\| \\
&\quad + (1-\alpha_{n+1})\Big(\|x_n - \overline{x}_n\| + \|x_n - T\overline{x}_n\|\Big) \\
&= \alpha_{n+1}\xi_n + (1-\alpha_{n+1})\Big(\|x_n - \overline{x}_n\| + \|x_n - T\overline{x}_n\|\Big).
\end{aligned} \tag{3.48}$$

On cherche à majorer les suites $(\|x_n - \overline{x}_n\|)_{n\in\mathbb{N}}$ et $(\|x_n - T\overline{x}_n\|)_{n\in\mathbb{N}}$. Pour tout entier $n \geqslant 2$, on a d'une part

$$\delta_n \leqslant \|\overline{x}_{n-1} - x_{n-1}\| + \lambda_{n-1}\xi_{n-1} = \lambda_{n-1}\xi_{n-1} + (1-\alpha_{n-1})\delta_{n-1} \tag{3.49}$$

et

$$\|x_n - \overline{x}_n\| \leqslant (1-\alpha_n)(\lambda_{n-1}\xi_{n-1} + (1-\alpha_{n-1})\delta_{n-1}), \tag{3.50}$$

et d'autre part

$$\begin{aligned}
\|\overline{x}_n - \overline{x}_{n-1}\| &\leqslant \alpha_n\|x_n - \overline{x}_{n-1}\| + (1-\alpha_n)\|x_{n-1} - \overline{x}_{n-1}\| \\
&\leqslant \alpha_n\lambda_{n-1}\xi_{n-1} + (1-\alpha_n)(1-\alpha_{n-1})(\lambda_{n-2}\xi_{n-2} + (1-\alpha_{n-2})\delta_{n-2})
\end{aligned} \tag{3.51}$$

et

$$\begin{aligned}
\|T\overline{x}_n - x_n\| &\leqslant \|T\overline{x}_n - T\overline{x}_{n-1}\| + \|T\overline{x}_{n-1} - x_n\| \\
&\leqslant \alpha_n\lambda_{n-1}\xi_{n-1} + (1-\alpha_n)(1-\alpha_{n-1})(\lambda_{n-2}\xi_{n-2} + (1-\alpha_{n-2})\delta_{n-2}) \\
&\quad + (1-\lambda_{n-1})\xi_{n-1}.
\end{aligned} \tag{3.52}$$

Ainsi, pour tout entier $n \geqslant 2$,

$$\begin{aligned}
\|T\overline{x}_n - x_n\| + \|x_n - \overline{x}_n\| \leqslant \xi_{n-1} + (1-\alpha_n)(1-\alpha_{n-1})(\delta_{n-1} + (1-\alpha_{n-2})\delta_{n-2} \\
+ \lambda_{n-2}\xi_{n-2})
\end{aligned} \tag{3.53}$$

et (3.48) devient alors

$$\begin{aligned}\xi_{n+1} &\leqslant \alpha_{n+1}\xi_n + (1-\alpha_{n+1})\xi_{n-1}\\ &\quad + (1-\alpha_{n+1})(1-\alpha_n)(1-\alpha_{n-1})(\delta_{n-1} + (1-\alpha_{n-2})\delta_{n-2} + \lambda_{n-2}\xi_{n-2}).\end{aligned} \tag{3.54}$$

Finalement, il est clair que

$$(\forall n \in \mathbb{N}) \quad \|x_{n+1} - x\| \leqslant \sum \mu_{n,j}\|x_j - x\|, \tag{3.55}$$

et ainsi [5, Lemma 2.3] assure que $(x_n)_{n\in\mathbb{N}}$ et $(\overline{x}_n)_{n\in\mathbb{N}}$ sont bornées, et par suite, $(\xi_n)_{n\in\mathbb{N}}$ et $(\delta_n)_{n\in\mathbb{N}}$ aussi. On sait que $\sum_{n\geqslant 2}|(1-\alpha_{n+1})(1-\alpha_n)(1-\alpha_{n-1})(\delta_{n-1} + (1-\alpha_{n-2})\delta_{n-2} + \lambda_{n-2}\xi_{n-2})| < +\infty$, et que $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est concentrante, donc $(\xi_n)_{n\in\mathbb{N}}$ converge et on conclut par (a). □

Remarque 3.27 Dans le cas particulier où $(\forall n \in \mathbb{N})$ $\alpha_n = 1 - 1/(n+1)$, la suite $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{Fix} T$.

Il a été démontré que certains algorithmes inertiels peuvent garantir un taux optimal, en un certain sens, sur les valeurs de la fonction objectif [2]. Nous nous demandons ici s'il est possible de conserver ce taux avec des algorithmes inertiels à 3 pas de mémoire.

Conjecture 3.28 *Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice d'ordre 3. Soit $f \in \Gamma_0(\mathcal{H})$ et soit $g\colon \mathcal{H} \to \mathbb{R}$ une fonction convexe différentiable avec un gradient $1/\beta$-lipschitzien. Supposons que* $\operatorname{Argmin}(f+g) \neq \varnothing$. *Soit $\gamma \in \left]0, \beta\right]$ et soit $x_{-2} = x_{-1} = x_0 \in \mathcal{H}$. De plus, soit $a \geqslant 2$ et on pose pour tout n entier strictement positif $\nu_n = (n-1)/(n+a)$. On itère*

$$\begin{array}{l}\textit{pour } n = 0, 1, \ldots\\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\ x_{n+1} = \operatorname{prox}_{\gamma f}(\overline{x}_n - \gamma\nabla g(\overline{x}_n)). \end{array}\right.\end{array} \tag{3.56}$$

Supposons que les conditions suivantes sont satisfaites :

(i) *Il existe deux suites $(\eta_n)_{n\in\mathbb{N}}$ et $(\sigma_n)_{n\in\mathbb{N}}$ à valeurs dans $\left]0,1\right[$ telles que*

$$(\forall n \in \{2,3,\ldots\})(\forall j \in \mathbb{N}) \quad \mu_{n,j} = \begin{cases} 1+\eta_n, & \textit{si } j = n;\\ \sigma_n - \eta_n, & \textit{si } j = n-1;\\ -\sigma_n, & \textit{si } j = n-2;\\ 0, & \textit{si } j < n-2.\end{cases} \tag{3.57}$$

(ii) $(\forall n \in \mathbb{N})$ $\eta_n \leqslant \nu_n$ *et* $\eta_{n+1}\nu_{n+1} + \sigma_{n+1} \leqslant \nu_{n+1}\nu_{n+2}$.

Alors les propriétés suivantes sont satisfaites :

(i) $(n\|x_n - x_{n-1}\|^2)_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$.

(ii) $(x_n)_{n\in\mathbb{N}}$ *converge faiblement.*

(iii) $((f+g)(x_n) - \min(f+g)) = O(1/n^2)$.

Les trois exemples qui suivent sont des applications directes du Théorème 3.23. Ils sont des généralisations respectivement de [1, Proposition 28.13], [1, Proposition 28.1(i)&(iii)] et [4, Theorem 7.3(i)]. Ces exemples montrent à nouveau que les matrices superconcentrantes permettent de généraliser aisément de nombreux résultats.

Proposition 3.29 *Soit $f \in \Gamma_0(\mathcal{H})$, soit $\beta \in \,]0,+\infty[$, soit $g\colon \mathcal{H} \to \mathbb{R}$ une fonction convexe différentiable avec un gradient $1/\beta$-lipschitzien, soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice superconcentrante de réels positifs, soit $\varepsilon \in \,]0,\min\{1,\beta\}[$, soit $(\gamma_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $[\varepsilon, 2\beta-\varepsilon]$, soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $[\varepsilon,1]$ et soit $x_0 \in \operatorname{dom} f$. Supposons que $\operatorname{Argmin}(f+g) \neq \varnothing$. On itère*

$$
\begin{array}{l}
\text{pour } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
\overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\
y_n = \overline{x}_n - \gamma_n \nabla g(\overline{x}_n) \\
z_n = \operatorname{prox}_{\gamma_n f} y_n \\
x_{n+1} = \overline{x}_n + \lambda_n\big(z_n - \overline{x}_n\big).
\end{array}
\right.
\end{array}
\tag{3.58}
$$

Alors les propriétés suivantes sont satisfaites :

(i) *Soit $x \in \operatorname{Argmin}(f+g)$. Alors $\sum_{n\in\mathbb{N}} \|\nabla g(\overline{x}_n) - \nabla g(x)\|^2 < +\infty$.*

(ii) $\sum_{n\in\mathbb{N}} \|z_n - \overline{x}_n\|^2 < +\infty$.

(iii) $\sum_{n\in\mathbb{N}}((f+g)(\overline{x}_n) - \inf(f+g)(\mathcal{H}))^2 < +\infty$.

(iv) $(\forall n \in \mathbb{N})\ (f+g)(x_{n+1}) \leqslant \sum_{j=0}^{n} \mu_{n,j} f(x_j)$.

(v) *$(x_n)_{n\in\mathbb{N}}$ et $(\overline{x}_n)_{n\in\mathbb{N}}$ convergent faiblement vers un même minimiseur de $f+g$.*

(vi) *Supposons que $f+g$ est uniformément quasiconvexe sur tout sous-ensemble borné de $\operatorname{dom} f$. Alors $(x_n)_{n\in\mathbb{N}}$ converge fortement vers l'unique minimiseur de $f+g$.*

Démonstration. Soit $n \in \mathbb{N}$ et $x \in \operatorname{Argmin}(f+g)$. On a

$$
\begin{aligned}
\|z_n - x\|^2 &= \|\operatorname{prox}_{\gamma_n f}(\overline{x}_n - \gamma_n \nabla g(\overline{x}_n)) - \operatorname{prox}_{\gamma_n f}(x - \gamma_n \nabla g(x))\|^2 \\
&\leqslant \|(\mathrm{Id} - \gamma_n \nabla g)\overline{x}_n - (\mathrm{Id} - \gamma_n \nabla g)x\|^2 - \|y_n - z_n + \gamma_n \nabla g(x)\|^2 \\
&\leqslant \|\overline{x}_n - x\|^2 - \gamma_n(2\beta - \gamma_n)\|\nabla g(\overline{x}_n) - \nabla g(x)\|^2 - \|y_n - z_n + \gamma_n \nabla g(x)\|^2
\end{aligned}
\tag{3.59}
$$

D'autre part, par la convexité de $\|\cdot\|^2$, on a $\|x_{n+1} - x\|^2 \leqslant (1-\lambda_n)\|\overline{x}_n - x\|^2 + \lambda_n \|z_n - x\|^2$.

Ainsi on a

$$\begin{aligned}\|x_{n+1}-x\|^2 &\leqslant \|\overline{x}_n - x\|^2 - \varepsilon\gamma_n(2\beta-\gamma_n)\|\nabla g(\overline{x}_n)-\nabla g(x)\|^2\\ &\quad - \varepsilon\|y_n - z_n + \gamma_n\nabla g(x)\|2 \qquad (3.60)\\ &\leqslant \|\overline{x}_n - x\|^2 - \varepsilon^3\|\nabla g(\overline{x}_n)-\nabla g(x)\|^2 \qquad (3.61)\\ &\leqslant \sum_{j=0}^{n}\mu_{n,j}\|x_j - x\|^2 \qquad (3.62)\end{aligned}$$

qui implique que $(\|x_k - x\|)_{k\in\mathbb{N}}$ converge.

(i) : Clair par (3.61) et superconcentration.

(ii) : Par (3.60), on sait que $\sum_{n\in\mathbb{N}}\|z_n - y_n - \gamma_n\nabla g(x)\|^2 < +\infty$. On conclut car $(z_n - x_n)_{n\in\mathbb{N}} = (z_n - y_n - \gamma_n\nabla g(x))_{n\in\mathbb{N}} + (\gamma_n\nabla g(x) - \gamma_n\nabla g(\overline{x}_n))_{n\in\mathbb{N}}$, et ces deux termes sont dans $\ell^2(\mathbb{N})$.

(iii)(iv) : Soit $z \in \mathcal{H}$ et soit $n \in \mathbb{N}$. Car $z_n = \operatorname{prox}_{\gamma_n f} y_n$, on déduit de [8, Proposition 12.26] que

$$\begin{aligned}\left\langle z - z_n \mid \gamma_n^{-1}(\overline{x}_n - z_n) - \nabla g(\overline{x}_n)\right\rangle &= \left\langle z - z_n \mid \gamma_n^{-1}(y_n - z_n)\right\rangle\\ &\leqslant f(z) - f(z_n). \qquad (3.63)\end{aligned}$$

D'autre part, on a

$$\langle z - z_n \mid \nabla g(z_n)\rangle + g(z_n) \leqslant g(z) \qquad (3.64)$$

par la [8, Proposition 17.7]. Soit $\mu = \sup_{k\in\mathbb{N}}\|z_k - z\|$. Car $(\|x_k - x\|)_{k\in\mathbb{N}}$ converge, (ii) assure que $\mu < +\infty$. De plus, avec (3.63) et (3.64), on obtient

$$\begin{aligned}&-\mu\|\overline{x}_n - z_n\|(1/\varepsilon + 1/\beta) + h(z_n)\\ &\quad\leqslant -\gamma_n^{-1}\|z_n - z\|\|\overline{x}_n - z_n\| - \|z_n - z\|\|\nabla g(\overline{x}_n) - \nabla g(z_n)\| + h(z_n)\\ &\quad\leqslant \left\langle z - z_n \mid \gamma_n^{-1}(x_n - z_n) - (\nabla g(x_n) - \nabla g(z_n))\right\rangle + h(z_n)\\ &\quad\leqslant h(z). \qquad (3.65)\end{aligned}$$

Ainsi $h(z_n) - h(z) \leqslant \mu(1/\varepsilon + 1/\beta)\|z_n - \overline{x}_n\| \leqslant (2\mu/\varepsilon)\|z_n - \overline{x}_n\|$ et si $z \in \operatorname{Argmin} h$, alors

$$0 \leqslant h(z_n) - \inf h(\mathcal{H}) \leqslant \frac{2\mu}{\varepsilon}\|z_n - x_n\|. \qquad (3.66)$$

On déduit de (3.63) que

$$f(z_n) \leqslant f(\overline{x}_n) - \left\langle \overline{x}_n - z_n \mid \gamma_n^{-1}(\overline{x}_n - z_n) - \nabla g(\overline{x}_n)\right\rangle \qquad (3.67)$$

et par le lemme de descente ([8, Theorem 18.15]), on a aussi

$$g(z_n) \leqslant g(\overline{x}_n) - \langle \overline{x}_n - z_n \mid \nabla g(\overline{x}_n)\rangle + \frac{1}{2\beta}\|z_n - \overline{x}_n\|^2. \qquad (3.68)$$

En additionnant (3.67) et (3.68), on obtient alors

$$h(z_n) \leqslant h(\overline{x}_n) - (\frac{1}{\gamma_n} - \frac{1}{2\beta})\|z_n - \overline{x}_n\|^2 \leqslant h(\overline{x}_n). \tag{3.69}$$

Donc, par convexité de h

$$\begin{aligned} h(x_{n+1}) &= h(\lambda_n z_n + (1-\lambda_n)\overline{x}_n) \\ &\leqslant \lambda_n h(z_n) + (1-\lambda_n)h(\overline{x}_n) \\ &= h(\overline{x}_n) - \lambda_n\big(h(\overline{x}_n) - h(z_n)\big) \\ &\leqslant h(\overline{x}_n) \end{aligned} \tag{3.70}$$

Ce qui montre que la relation requise par (iv) est bien vérifiée et que $\sum_{k\in\mathbb{N}} \lambda_k(h(\overline{x}_k) - h(z_k)) < +\infty$. Car $\inf_{k\in\mathbb{N}} \lambda_k \geqslant \varepsilon$,

$$\sum_{k\in\mathbb{N}} h(\overline{x}_k) - h(z_k) < +\infty. \tag{3.71}$$

Ainsi, en utilisant (ii) et (3.66), on a $(h(z_k) - \inf h(\mathcal{H}))_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$. Ainsi, $(h(\overline{x}_k) - \inf h(\mathcal{H}))_{k\in\mathbb{N}} \in \ell^2(\mathbb{N})$.

(v) : Puisque $(\|x_n - x\|)_{n\in\mathbb{N}}$ converge, on doit montrer avec $x_{k_n} \rightharpoonup x \in \mathcal{H}$, que $x \in \operatorname{Argmin}(f+g)$. Car $\inf_{k\in\mathbb{N}} \lambda_k \geqslant \varepsilon$, on a $x_n - \overline{x}_n \to 0$ et donc $\overline{x}_{k_n} \rightharpoonup x$. Or, puisque $h(\overline{x}_{k_n}) \to \min h(\mathcal{H})$ et $h \in \Gamma_0(\mathcal{H})$, on conclut que $x \in \operatorname{Argmin}(f+g)$.

(vi) : On sait par la [1, Proposition 11.8(i)] que h a un unique minimiseur, on l'appelle x. Puisque $h(\overline{x}_n) \to \min h(\mathcal{H})$, on pose $C = \{x_n\}_{n\in\mathbb{N}} \cup \{x\} \subset \operatorname{dom} h$. Soit ϕ le module d'uniforme de quasiconvexité de h sur C, et soit $\alpha \in]0,1[$. Par définition de ϕ, pour tout $n \in \mathbb{N}$, on a

$$\begin{aligned} h(x) + \alpha(1-\alpha)\phi(\|\overline{x}_n - x\|) &= \inf h(\mathcal{H}) + \alpha(1-\alpha)\phi(\|\overline{x}_n - x\|) \\ &= h(\alpha\overline{x}_n + (1-\alpha)x) + alpha(1-\alpha)\phi(\|\overline{x}_n - x\|) \\ &= \max\{h(\overline{x}_n, h(x)\} \\ &= h(\overline{x}_n). \end{aligned} \tag{3.72}$$

On sait que $h(\overline{x}_n) \to h(x)$, donc $\phi(\|\overline{x}_n - x\|) \to 0$ et ainsi on peut conclure car $\|\overline{x}_n - x\| \to 0$ et $\overline{x}_n - x_n \to 0$. □

Proposition 3.30 *Soit $f \in \Gamma_0(\mathcal{H})$ telle que* Argmin $f \neq \varnothing$*, soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice superconcentrante de réels positifs, soit $(\gamma_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,+\infty[$ et soit $(\lambda_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $]0,1[$ telle que $\sum_{n\in\mathbb{N}} \lambda_n\gamma_n = +\infty$ et $\sum_{n\in\mathbb{N}} \lambda_n = +\infty$. Soit $x_0 \in \mathcal{H}$ et on itère*

$$\begin{array}{l} \text{pour } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = \sum_{j=0}^{n} \mu_{n,j} x_j \\ y_n = \operatorname{prox}_{\gamma_n f} \overline{x}_n \\ x_{n+1} = \overline{x}_n + \lambda_n\big(y_n - \overline{x}_n\big). \end{array} \right. \end{array} \tag{3.73}$$

Alors les propriétés suivantes sont satisfaites :

(i) $\sum_{n\in\mathbb{N}} \lambda_n\gamma_n(f(y_n) - \min f(\mathcal{H})) < +\infty$ *et* $\sum_{n\in\mathbb{N}} \lambda_n(f(y_n) - f(\overline{x}_n)) < +\infty$

(ii) $(\forall n \in \mathbb{N})\ f(x_{n+1}) \leqslant \sum_{j=0}^{n} \mu_{n,j} f(x_j)$.

(iii) $f(x_n) \to \min f(\mathcal{H})$

(iv) $(x_n)_{n\in\mathbb{N}}$ *converge faiblement vers un minimiseur de* f.

Démonstration. Par (3.73) et par la définition de prox_f, on a

$$(\forall n \in \mathbb{N}) \quad \overline{x}_n - y_n \in \gamma_n \partial f(y_n) \tag{3.74}$$

et la définition de ∂f montre que

$$(\forall z \in \mathcal{H})(\forall n \in \mathbb{N}) \quad \langle z - y_n \mid \overline{x}_n - y_n\rangle \leqslant \gamma_n(f(z) - f(y_n)). \tag{3.75}$$

On obtient de (3.74) que

$$(\forall n \in \mathbb{N}) \quad 0 \leqslant \langle \overline{x}_n - y_n \mid \overline{x}_n - y_n\rangle/\gamma_n \leqslant f(\overline{x}_n) - f(y_n). \tag{3.76}$$

Soit $z \in \operatorname{Argmin} f$. La convexité de $\|\cdot\|^2$ assure que pour tout $n \in \mathbb{N}$, $\|x_{n+1} - x\|^2 \leqslant (1-\lambda_n)\|\overline{x}_n - z\|^2 + \lambda_n\|y_n - z\|^2$ et que

$$\begin{aligned}
\|y_n - z\|^2 &= \|\overline{x}_n - z\|^2 + 2\langle \overline{x}_n - z \mid y_n - \overline{x}_n\rangle + \|y_n - \overline{x}_n\|^2 \\
&= \|\overline{x}_n - z\|^2 - \|y_n - \overline{x}_n\|^2 + 2\langle y_n - z \mid y_n - \overline{x}_n\rangle \\
&\leqslant \|\overline{x}_n - z\|^2 - 2\gamma_n(f(y_n) - \min f(\mathcal{H})).
\end{aligned} \tag{3.77}$$

Car $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante, on a $\sum_{n\in\mathbb{N}} \lambda_n\gamma_n(f(y_n) - \min f(\mathcal{H})) < +\infty$. Ainsi, car $\sum_{n\in\mathbb{N}} \lambda_n\gamma_n = +\infty$, on a $\underline{\lim}\, f(y_n) = \min f(\mathcal{H})$.

(ii) : Il suit de (3.76) que

$$(\forall n \in \mathbb{N}) \quad f(x_{n+1}) \leqslant (1-\lambda_n)f(\overline{x}_n) + \lambda_n f(y_n) = f(\overline{x}_n) + \lambda_n(f(y_n) - f(\overline{x}_n)) \leqslant \sum_{i=0}^{n} \mu_{n,i} f(x_i). \tag{3.78}$$

Donc on a $\sum_{n\in\mathbb{N}} \lambda_n(f(y_n) - f(\overline{x}_n)) < +\infty$.

(iii) : Par (ii), on a $\underline{\lim}\, f(y_n) = \min f(\mathcal{H})$, $\underline{\lim}\, f(y_n) - f(\overline{x}_n) = 0$, et puisque $(f(\overline{x}_n))_{n\in\mathbb{N}}$ converge, on déduit que $f(\overline{x}_n) \to \min f(\mathcal{H})$ et $f(y_n) \to \min f(\mathcal{H})$. De plus, par (3.78), on a $f(x_n) \to \min f(\mathcal{H})$.

(iv) : Soit x un point d'accumulation séquentiel faible de $(x_n)_{n\in\mathbb{N}}$. On déduit de (ii) et de la [8, Proposition 11.21] que $x \in \operatorname{Argmin} f$. Il suffit d'appliquer le [8, Theorem 5.5]. ☐

Proposition 3.31 *Soit $f\colon \mathcal{H} \to \mathbb{R}$ une fonction convexe continue et soit C un sous-ensemble convexe fermé de $\mathcal{H}$. Le problème est de*

$$\text{trouver} \quad x \in C \quad \text{tel que} \quad f(x) = \overline{f} \quad \text{où} \quad \overline{f} = \inf f(C) \tag{3.79}$$

Appelons S l'ensemble des solutions de ce problème et supposons le non vide. Supposons que l'ensemble des solutions de ce problème est non vide et on l'appelle S. Soit $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ une matrice superconcentrante de réels positifs, soit $\delta \in]0,1[$, soit $(\alpha_n)_{n\in\mathbb{N}} \in \ell^\infty(\mathbb{N}) \smallsetminus \ell^1(\mathbb{N})$, soit $(\varepsilon_n)_{n\in\mathbb{N}} \in \ell^1(\mathbb{N})$, soit $(\nu_n)_{n\in\mathbb{N}}$ à valeurs dans $]\delta, 2-\delta[$ et soit $(s_n)_{n\in\mathbb{N}}$ une sélection de $(\partial_{\varepsilon_n} f)_{n\in\mathbb{N}}$. Soit $x_0 \in \mathcal{H}$. On itère

$$\begin{array}{l} \textit{for } n = 0,1,\ldots \\ \left\lfloor \begin{array}{l} \overline{x}_n = \displaystyle\sum_{j=0}^{n} \mu_{n,j} x_j \\ y_n = \overline{x}_n - \alpha_n s_n(\overline{x}_n) \\ x_{n+1} = y_n + \nu_n(P_C y_n - y_n) \end{array} \right. \end{array} \tag{3.80}$$

Supposons qu'il existe $\kappa \in]0,+\infty[$, $(\beta_n)_{n\in\mathbb{N}} \in \ell_+(\mathbb{N}) \cap \ell^1(\mathbb{N})$, et $(\gamma_n)_{n\in\mathbb{N}} \in \ell_+(\mathbb{N}) \cap \ell^1(\mathbb{N})$ tels que

$$(\forall n \in \mathbb{N}) \quad \overline{f} - \gamma_{n+1} \leqslant f(x_{n+1}) \leqslant f(\overline{x}_n) - \kappa\alpha_n \|u_n\|^2 + \beta_n \tag{3.81}$$

Alors $(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un minimiseur de f et $\lim f(\overline{x}_n) = \min f(C)$.

Démonstration. On pose $\zeta = \overline{f} - \sup_{n\in\mathbb{N}} \gamma_n$. Par (3.81), on a

$$(\forall n \in \mathbb{N}) \quad 0 \leqslant f(x_{n+1}) - \zeta \leqslant (f(\overline{x}_n) - \zeta) - \kappa\alpha_n \|u_n\|^2 + \beta_n. \tag{3.82}$$

Car $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ est superconcentrante, on a la convergence de $(f(x_n))_{n\in\mathbb{N}}$ et $(\alpha_n\|u_n\|^2) \in \ell^1(\mathbb{N})$. Donc, pour $n \in \mathbb{N} \smallsetminus \{0\}$, on a $\overline{f} - f(\overline{x}_n) \leqslant \gamma_n$ et

$$\sum_{k\in\mathbb{N}} \alpha_k \|u_k\|^2 + 2(\overline{f} - f(\overline{x}_k))^+ + 2\varepsilon_k \leqslant \sum_{k\in\mathbb{N}} \alpha_k \|u_k\|^2 + 2\gamma_k + 2\varepsilon_k < +\infty \tag{3.83}$$

Maintenant, on pose $(\forall n \in \mathbb{N})$ $T_{1,n} = \mathrm{Id} + \nu_n(P_C - \mathrm{Id})$ et $T_{2,n} = \mathrm{Id} - \alpha_n s_n$. Ainsi

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = T_{1,n} T_{2,n} \overline{x}_n \tag{3.84}$$

Soit $x \in \operatorname{Argmin} f$. Car P_C est une contraction ferme, on a, pour tout $n \in \mathbb{N}$,

$$\begin{aligned} \|x_{n+1} - x\|^2 \leqslant& \|\overline{x}_n - x\|^2 - \nu_n(2-\nu_n) d_C(y_n)^2 \\ &+ \alpha_n^2 \|u_n\|^2 + 2\alpha_n(\overline{f} - f(\overline{x}_n)) + 2\alpha_n\varepsilon_n. \end{aligned} \tag{3.85}$$

La superconcentration de $(\mu_{n,j})_{n\in\mathbb{N},0\leqslant j\leqslant n}$ assure que

$$\sum_{n\in\mathbb{N}} \alpha_n (2(f(\overline{x}_n) - \overline{f}) - 2\varepsilon_n - \alpha_n \|u_n\|^2) < +\infty. \tag{3.86}$$

On a $\sum_{n\in\mathbb{N}} \alpha_n = +\infty$ et donc

$$\underline{\lim}\, 2(f(\overline{x}_n) - \overline{f}) - 2\varepsilon_n - \alpha_n \|u_n\|^2 \leqslant 0. \tag{3.87}$$

On conclut que $\underline{\lim}\, f(\overline{x}_n) \leqslant \overline{f}$. Pour utiliser le Théorème 3.23(iv), il reste à montrer que tout point d'accumulation séquentiel faible de $(\overline{x}_n)_{n\in\mathbb{N}}$ est dans S. Soit une sous-suite de $(\overline{x}_n)_{n\in\mathbb{N}}$ telle que $x_{k_n} \rightharpoonup x$. L'hypothèse sur $(\nu_n)_{n\in\mathbb{N}}$ et (3.85) impliquent que $d_C(y_n) \to 0$. Sachant que $(\alpha_n)_{n\in\mathbb{N}} \in l^\infty$, on a $\alpha_n\|u_n\| \to 0$ et ainsi $\overline{x}_{k_n} - \alpha_{k_n} u_{k_n} \rightharpoonup x$. Par faible semicontinuité inférieure de d_C, on déduit que $d_C(x) \leqslant \underline{\lim}\, d_C(\overline{x}_{k_n} - \alpha_{k_n} u_{k_n}) = 0$. On a donc $x \in C$ et, ainsi $\overline{f} \leqslant f(x)$. Puisque f est faiblement semicontinue inférieurement,

$$\overline{f} \leqslant f(x) \leqslant \underline{\lim}\, f(\overline{x}_{k_n}) = \lim f(\overline{x}_n) \leqslant \overline{f}. \tag{3.88}$$

On a ainsi $f(x) = \overline{f}$ et $x \in C$, et donc $x \in S$, ce qui conclut la démonstration. □

3.5 Bibliographie

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. Springer, New York, 2017.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems,* SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[3] A. Chambolle and C. Dossal, *On the convergence of the iterates of the "Fast iterative shrinkage/thresholding algorithm",* J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[4] P. L. Combettes, *Quasi-Fejérian analysis of some optimization algorithms,* in : Inherently Parallel Algorithms for Feasibility and Optimization, D. Butnariu, Y. Censor, and S. Reich, eds. Elsevier, New York, 2001, pp. 115–152.

[5] P. L. Combettes and T. Pennanen, *Generalized Mann iterates for constructing fixed points in Hilbert spaces,* J. Math. Anal. Appl., 275 (2002), pp. 521–536.

[6] P. L. Combettes and I. Yamada, *Compositions and convex combinations of averaged nonexpansive operators,* J. Math. Anal. Appl., 425 (2015), pp. 55–70.

[7] R. Cominetti, J. A. Soto, and J. Vaisman, *On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis,* Israel J. Math., 199 (2014), pp. 757–772.

[8] K. Knopp, *Theory and Application of Infinite Series,* Blackie & Son, London, 1954.

[9] H. F. Walker and P. Ni, *Anderson acceleration for fixed-point iterations,* SIAM J. Numer. Anal., 49 (2011), pp. 1715–1735.

Chapitre 4

Itérations moyennées à métrique variable

L'intérêt d'une métrique variable dans certains algorithmes d'optimisation est bien connu, cf. [16, 22] et leur bibliographie. Dans ce chapitre nous étendons cette stratégie au cadre des méthodes de point fixe multicouches avec opérateurs moyennés. Pour compléter cette étude nous fournissons des applications de cet algorithme, en particulier dans le cas d'un problème d'inclusion monotone composite.

4.1 Description et résultats principaux

Nous allons étudier l'inclusion monotone fondamentale qui suit.

Problème 4.1 Soit A un opérateur maximalement monotone, soit $\beta \in \left]0,+\infty\right[$ et soit $B : \mathcal{H} \to \mathcal{H}$ un opérateur β-cocoercif. De plus, on suppose que

$$S = \text{zer}\,(A+B) \neq \varnothing. \tag{4.1}$$

L'objectif est de trouver un point dans S.

Nous utilisons le cadre général des opérateurs moyennés. On rappelle que l'on cherche à produire un algorithme pour lequel la métrique peut changer à chaque itération. Pour cela, nous allons avoir besoin de la définition suivante.

Définition 4.2 Soit $\mu \in \left]0,+\infty\right[$, soit $U \in \mathcal{P}_\mu(\mathcal{H})$, soit $\alpha \in \left]0,1\right]$ et soit $T\colon \mathcal{H} \to \mathcal{H}$ un opérateur. Alors T est α-moyenné sur $(\mathcal{H}, U)$ si

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx - Ty\|_U^2 \leqslant \|x-y\|_U^2 - \frac{1-\alpha}{\alpha}\|Tx - x\|_U^2. \tag{4.2}$$

Si $\alpha = 1$, on dit que T est contractant sur $(\mathcal{H}, U)$.

Nous allons résoudre le problème auxiliaire suivant.

Problème 4.3 Soit $\alpha \in \,]0, +\infty[$, soit $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ et soit $(U_n)_{n\in\mathbb{N}}$ une suite de $\mathcal{P}_\alpha(\mathcal{H})$ telle que

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{et} \quad (\forall n \in \mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.3}$$

Soit $m \geqslant 1$ un entier. Pour tout $n \in \mathbb{N}$ et tout $i \in \{1, \ldots, m\}$, soit $\alpha_{i,n} \in \,]0, 1]$ et soit $T_{i,n} \colon \mathcal{H} \to \mathcal{H}$ $\alpha_{i,n}$-moyenné sur $(\mathcal{H}, U_n^{-1})$. L'objectif est de

$$\text{trouver} \quad x \in S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix} T_n \neq \varnothing, \quad \text{où} \quad (\forall n \in \mathbb{N}) \quad T_n = T_{1,n} \cdots T_{m,n}. \tag{4.4}$$

On présente maintenant le résultat central de ce chapitre.

Théorème 4.4 *On considère le cadre du Problème 4.3. Soit $\varepsilon \in \,]0, 1[$ et pour tout $n \in \mathbb{N}$, soit ϕ_n une constante de moyennage de T_n, soit $\lambda_n \in \,]0, \phi_n]$ et soit $e_{i,n} \in \mathcal{H}$. On itère*

$$\begin{array}{l} \textit{pour } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}x_n + e_{m,n}) + e_{m-1,n} \cdots\big) + e_{2,n}\Big) + e_{1,n} \\ x_{n+1} = x_n + \lambda_n(y_n - x_n). \end{array} \right. \end{array} \tag{4.5}$$

Supposons que

$$(\forall i \in \{1, \ldots, m\}) \quad \sum_{n\in\mathbb{N}} \lambda_n \|e_{i,n}\|_{U_n^{-1}} < +\infty, \tag{4.6}$$

et on définit

$$(\forall i \in \{1, \ldots, m\})(\forall n \in \mathbb{N}) \quad T_{i+,n} = \begin{cases} T_{i+1,n} \cdots T_{m,n}, & \textit{if } i \neq m; \\ \mathrm{Id}\,, & \textit{if } i = m. \end{cases} \tag{4.7}$$

Alors les propriétés suivantes sont satisfaites :

(i) $\sum_{n\in\mathbb{N}} \lambda_n(1/\phi_n - \lambda_n)\|T_{1,n} \cdots T_{m,n}x_n - x_n\|^2_{U_n^{-1}} < +\infty$.

(ii) *Supposons que* $(\forall n \in \mathbb{N})$ $\lambda_n \in \,]0, \varepsilon + (1-\varepsilon)/\phi_n]$. *Alors* $(\forall x \in S)$

$$\max_{1\leqslant i\leqslant m} \sum_{n\in\mathbb{N}} \frac{\lambda_n(1-\alpha_{i,n})}{\alpha_{i,n}} \|(\mathrm{Id} - T_{i,n})T_{i+,n}x_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2_{U_n^{-1}} < +\infty. \tag{4.8}$$

(iii) $(x_n)_{n\in\mathbb{N}}$ *converge faiblement vers un point de S si et seulement si tout point d'accumulation séquentiel faible de $(x_n)_{n\in\mathbb{N}}$ est dans S. Dans ce cas, la convergence est forte si* $\operatorname{int} S \neq \varnothing$.

(iv) $(x_n)_{n\in\mathbb{N}}$ *converge fortement vers un point de S si et seulement si* $\underline{\lim}\, d_S(x_n) = 0$.

On résout le Problème 4.3 grâce à un algorithme explicite-implicite sur-relaxé à métrique variable.

Proposition 4.5 *Soit $\beta \in \left]0,+\infty\right[$, soit $\varepsilon \in \left]0,\min\{1/2,\beta\}\right[$, soit $\alpha \in \left]0,+\infty\right[$, soit $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ et soit $(U_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathcal{P}_\alpha(\mathcal{H})$ telle que*

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{et} \quad (\forall n \in \mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.9}$$

Soit $x_0 \in \mathcal{H}$, soit $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ un opérateur maximalement monotone et soit $B\colon \mathcal{H} \to \mathcal{H}$ un opérateur β-cocoercif. Soient $(a_n)_{n\in\mathbb{N}}$ et $(b_n)_{n\in\mathbb{N}}$ deux suites à valeurs dans $\mathcal{H}$ telles que $\sum_{n\in\mathbb{N}} \|a_n\| < +\infty$ et $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$. Supposons que $\operatorname{zer}(A+B) \neq \varnothing$ et, pour tout $n \in \mathbb{N}$, soient

$$\gamma_n \in \left[\varepsilon, \frac{2\beta}{(1+\varepsilon)\|U_n\|}\right] \quad \text{et} \quad \lambda_n \in \left[\varepsilon, 1 + (1-\varepsilon)\left(1 - \frac{\gamma_n\|U_n\|}{2\beta}\right)\right]. \tag{4.10}$$

On itère

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n\Big(J_{\gamma_n U_n A}\big(x_n - \gamma_n U_n(Bx_n + b_n)\big) + a_n - x_n\Big). \tag{4.11}$$

Alors les propriétés suivantes sont satisfaites :

(i) $\sum_{n\in\mathbb{N}} \|J_{\gamma_n U_n A}(x_n - \gamma_n U_n B x_n) - x_n\|^2 < +\infty$.

(ii) *Soit $x \in \operatorname{zer}(A+B)$. Alors $\sum_{n\in\mathbb{N}} \|Bx_n - Bx\|^2 < +\infty$.*

(iii) *$(x_n)_{n\in\mathbb{N}}$ converge faiblement vers un point de $\operatorname{zer}(A+B)$.*

(iv) *Supposons que l'une des propriétés suivantes est satisfaite :*

  (a) *A est demi-régulier en tout point de $\operatorname{zer}(A+B)$.*

  (b) *B est demi-régulier en tout point de $\operatorname{zer}(A+B)$.*

  (c) $\operatorname{int} S \neq \varnothing$.

  Alors $(x_n)_{n\in\mathbb{N}}$ converge fortement vers un point de $\operatorname{zer}(A+B)$.

Maintenant, au lieu de décomposer l'opérateur explicite-implicite en deux opérateurs moyennés, considérons-le comme un seul opérateur. Nous ne pouvons plus travailler en dimension infinie, mais cela permet de proposer un second algorithme explicite-implicite dont les paramètres sont étroitement liés.

Exemple 4.6 Soit $\alpha \in \left]0,+\infty\right[$, soit $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$ et soit $(U_n)_{n\in\mathbb{N}}$ une suite à valeurs dans $\mathcal{P}_\alpha(\mathcal{H})$ telle que

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{et} \quad (\forall n \in \mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.12}$$

Soit $\varepsilon \in \,]0,1/2[$, soit $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ un opérateur maximalement monotone, soit $\beta \in \,]0,+\infty[$, soit B un opérateur β-cocoercif et soient $(\gamma_n)_{n\in\mathbb{N}}$ et $(\mu_n)_{n\in\mathbb{N}}$ deux suites à valeurs dans $[\varepsilon,+\infty[$ telles que

$$\phi_n = \frac{2\mu_n\beta}{4\beta - \|U_n\|\gamma_n} \leqslant 1-\varepsilon. \tag{4.13}$$

Soit $x_0 \in \mathcal{H}$ et on itère

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \mu_n\Big(J_{\gamma_n U_n A}(x_n - \gamma_n U_n B) - x_n\Big). \tag{4.14}$$

Supposons que $\mathcal{H}$ est de dimension finie et que $\operatorname{zer}(A+B) \neq \varnothing$. Alors $(x_n)_{n\in\mathbb{N}}$ converge vers un point de $\operatorname{zer}(A+B)$.

4.2 Article en anglais

VARIABLE METRIC ALGORITHMS DRIVEN BY AVERAGED OPERATORS [1]

Abstract. The convergence of a new general variable metric algorithm based on compositions of averaged operators is established. Applications to monotone operator splitting are presented.

Keywords. averaged operator, composite algorithm, convex optimization, fixed point iteration, monotone operator splitting, primal-dual algorithm, variable metric

AMS 2010 Subject Classification : 47H05, 49M27, 49M29, 90C25

4.2.1 Introduction

Iterations of averaged nonexpansive operators provide a synthetic framework for the analysis of many algorithms in nonlinear analysis, e.g., [3, 4, 7, 9, 18]. We establish the convergence of a new general variable metric algorithm based on compositions of averaged operators. These results are applied to the analysis of the convergence of a new forward-backward algorithm for solving the inclusion

$$0 \in Ax + Bx, \tag{4.15}$$

1. L. E. Glaudin, *Variable metric algorithms driven by averaged operators,* à paraître dans H. H. Bauschke, R. Burachick, and R. L. Luke (eds.), *Splitting Algorithms, Modern Operator Theory, and Applications,* Springer, New York.

where A and B are maximally monotone operators on a real Hilbert space. The theory of monotone operators is used in many applied mathematical fields, including optimization [10], partial differential equations and evolution inclusions [5, 21, 23], signal processing [13, 17], and statistics and machine learning [12, 19, 20]. In recent years, variants of the forward-backward algorithm with variable metric have been proposed in [15, 16, 22, 24], as well as variants involving overrelaxations [18]. The goal of the present paper is to unify these two approaches in the general context of iterations of compositions of averaged operators. In turn, this provides new methods to solve the problems studied in [1, 4, 6, 8, 9, 14].

The paper is organized as follows : Section 4.2.2 presents the background and notation. We establish the proof of the convergence of the general algorithm in Section 4.2.3. Special cases are provided in Section 4.2.4. Finally, by recasting these results in certain product spaces, we present and solve a general monotone inclusion in Section 4.2.5.

4.2.2 Notation and background

Throughout this paper, $\mathcal{H}$, $\mathcal{G}$, and $(\mathcal{G}_i)_{1\leqslant i\leqslant m}$ are real Hilbert spaces. We use $\langle\cdot\mid\cdot\rangle$ to denote the scalar product of a Hilbert space and $\|\cdot\|$ for the associated norm. Weak and strong convergence are respectively denoted by $\rightharpoonup$ and $\to$. We denote by $\mathcal{B}\,(\mathcal{H},\mathcal{G})$ the space of bounded linear operators from $\mathcal{H}$ to $\mathcal{G}$, and set $\mathcal{B}\,(\mathcal{H}) = \mathcal{B}\,(\mathcal{H},\mathcal{H})$ and $\mathcal{S}\,(\mathcal{H}) = \big\{L\in\mathcal{B}\,(\mathcal{H}) \;\big|\; L=L^*\big\}$, where L^* denotes the adjoint of L, and Id denotes the identity operator. The Loewner partial ordering on $\mathcal{S}\,(\mathcal{H})$ is defined by

$$(\forall U\in\mathcal{S}\,(\mathcal{H}))(\forall V\in\mathcal{S}\,(\mathcal{H}))\quad U\succcurlyeq V\quad\Leftrightarrow\quad(\forall x\in\mathcal{H})\quad\langle Ux\mid x\rangle\geqslant\langle Vx\mid x\rangle. \tag{4.16}$$

Let $\alpha\in\left]0,+\infty\right[$. We set

$$\mathcal{P}_\alpha(\mathcal{H}) = \big\{U\in\mathcal{S}\,(\mathcal{H}) \;\big|\; U\succcurlyeq\alpha\,\mathrm{Id}\,\big\}, \tag{4.17}$$

and we denote by $\sqrt{U}$ the square root of $U\in\mathcal{P}_\alpha(\mathcal{H})$. Moreover, for every $U\in\mathcal{P}_\alpha(\mathcal{H})$, we define a scalar product and a norm by

$$(\forall x\in\mathcal{H})(\forall y\in\mathcal{H})\quad\langle x\mid y\rangle_U=\langle Ux\mid y\rangle\quad\text{and}\quad\|x\|_U=\sqrt{\langle Ux\mid x\rangle}, \tag{4.18}$$

and we denote this Hilbert space by $(\mathcal{H},U)$. Let $A\colon\mathcal{H}\to2^{\mathcal{H}}$ be a set-valued operator. We denote by $\operatorname{dom}A=\big\{x\in\mathcal{H}\;\big|\;Ax\neq\varnothing\big\}$ the domain of A, by $\operatorname{gra}A=\big\{(x,u)\in\mathcal{H}\times\mathcal{H}\;\big|\;u\in Ax\big\}$ the graph of A, by $\operatorname{ran}A=\big\{u\in\mathcal{H}\;\big|\;(\exists\,x\in\mathcal{H})\;u\in Ax\big\}$ the range of A, by $\operatorname{zer}A=\big\{x\in\mathcal{H}\;\big|\;0\in Ax\big\}$ the set of zeros of A, and by A^{-1} the inverse of A which is the operator with graph $\big\{(u,x)\in\mathcal{H}\times\mathcal{H}\;\big|\;u\in Ax\big\}$. The resolvent of A is $J_A=(\mathrm{Id}\,+A)^{-1}$. Moreover, A is monotone if

$$(\forall(x,u)\in\operatorname{gra}A)(\forall(y,v)\in\operatorname{gra}A)\quad\langle x-y\mid u-v\rangle\geqslant0, \tag{4.19}$$

and maximally monotone if there exists no monotone operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ such that $\operatorname{gra} A \subset \operatorname{gra} B \neq \operatorname{gra} A$. The parallel sum of $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ and $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is

$$A \,\square\, B = (A^{-1} + B^{-1})^{-1}. \tag{4.20}$$

An operator $B\colon \mathcal{H} \to 2^{\mathcal{H}}$ is cocoercive with constant $\beta \in \left]0,+\infty\right[$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \langle x-y \mid Bx - By\rangle \geqslant \beta\|Bx-By\|^2. \tag{4.21}$$

Let C be a nonempty subset of $\mathcal{H}$. The interior of C is $\operatorname{int} C$. Finally, the set of summable sequences in $\left[0,+\infty\right[$ is denoted by $\ell_+^1(\mathbb{N})$.

Definition 4.7 Let $\mu \in \left]0,+\infty\right[$, let $U \in \mathcal{P}_\mu(\mathcal{H})$, let $\alpha \in \left]0,1\right]$, and let $T\colon \mathcal{H} \to \mathcal{H}$ be an operator. Then T is an α-averaged operator on $(\mathcal{H},U)$ if

$$(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|Tx-Ty\|_U^2 \leqslant \|x-y\|_U^2 - \frac{1-\alpha}{\alpha}\|Tx-x\|_U^2. \tag{4.22}$$

If $\alpha = 1$, T is nonexpansive on $(\mathcal{H},U)$.

Lemma 4.8 [4, Proposition 4.46] *Let $m \geqslant 1$ be an integer. For every $i \in \{1,\ldots,m\}$, let $T_i\colon \mathcal{H} \to \mathcal{H}$ be averaged. Then $T_1 \cdots T_m$ is averaged.*

Lemma 4.9 [4, Proposition 4.35] *Let $\mu \in \left]0,+\infty\right[$, let $U \in \mathcal{P}_\mu(\mathcal{H})$, let $\alpha \in \left]0,1\right]$, and let T be an α-averaged operator on $(\mathcal{H},U)$. Then the operator $R = (1-1/\alpha)\operatorname{Id} + (1/\alpha)T$ is nonexpansive on $(\mathcal{H},U)$.*

Lemma 4.10 [4, Lemma 5.31] *Let $(\alpha_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\left[0,+\infty\right[$, let $(\eta_n)_{n\in\mathbb{N}}$ and $(\varepsilon_n)_{n\in\mathbb{N}}$ be sequences in $\in \ell_+^1(\mathbb{N})$ such that*

$$(\forall n \in \mathbb{N}) \quad \alpha_{n+1} \leqslant (1+\eta_n)\alpha_n - \beta_n + \varepsilon_n. \tag{4.23}$$

Then $(\beta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$.

Lemma 4.11 [15, Proposition 4.1] *Let $\alpha \in \left]0,+\infty\right[$, let $(W_n)_{n\in\mathbb{N}}$ be in $\mathcal{P}_\alpha(\mathcal{H})$, let C be a nonempty subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that*

$$\begin{aligned}\big(\exists\, (\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)(\forall z \in C)\big(\exists\, (\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)(\forall n \in \mathbb{N})&\\ \|x_{n+1}-z\|_{W_{n+1}}^2 \leqslant (1+\eta_n)\|x_n - z\|_{W_n}^2 + \varepsilon_n.&\end{aligned} \tag{4.24}$$

Then $(x_n)_{n\in\mathbb{N}}$ is bounded and, for every $z \in C$, $(\|x_n - z\|_{W_n})_{n\in\mathbb{N}}$ converges.

Proposition 4.12 [15, Theorem 3.3] *Let $\alpha \in \left]0,+\infty\right[$, and let $(W_n)_{n\in\mathbb{N}}$ and W be operators in $\mathcal{P}_\alpha(\mathcal{H})$ such that $W_n \to W$ pointwise, as is the case when*

$$\sup_{n\in\mathbb{N}} \|W_n\| < +\infty \quad \text{and} \quad \big(\exists\, (\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)(\forall n \in \mathbb{N}) \quad (1+\eta_n)W_n \succcurlyeq W_{n+1}. \tag{4.25}$$

Let C be a nonempty subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that (4.24) is satisfied. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in C if and only if every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in C.

Proposition 4.13 [16, Proposition 3.6] *Let $\alpha \in \left]0,+\infty\right[$, let $(\nu_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(W_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that $\sup_{n\in\mathbb{N}} \|W_n\| < +\infty$ and $(\forall n \in \mathbb{N})$ $(1+\nu_n)W_{n+1} \succcurlyeq W_n$. Furthermore, let C be a subset of $\mathcal{H}$ such that $\operatorname{int} C \neq \varnothing$ and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that*

$$\begin{aligned}\big(\exists\,(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)\big(\exists\,(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)(\forall x \in \mathcal{H})(\forall n \in \mathbb{N})&\\ \|x_{n+1}-x\|_{W_{n+1}}^2 \leqslant (1+\eta_n)\|x_n - x\|_{W_n}^2 + \varepsilon_n.&\end{aligned} \tag{4.26}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges strongly.

Proposition 4.14 [15, Proposition 3.4] *Let $\alpha \in \left]0,+\infty\right[$, let $(W_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that $\sup_{n\in\mathbb{N}} \|W_n\| < +\infty$, let C be a nonempty closed subset of $\mathcal{H}$, and let $(x_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{H}$ such that*

$$\begin{aligned}\big(\exists\,(\varepsilon_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)\big(\exists\,(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})\big)(\forall z \in C)(\forall n \in \mathbb{N})&\\ \|x_{n+1}-z\|_{W_{n+1}}^2 \leqslant (1+\eta_n)\|x_n - z\|_{W_n}^2 + \varepsilon_n.&\end{aligned} \tag{4.27}$$

Then $(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in C if and only if $\underline{\lim}\, d_C(x_n) = 0$.

Lemma 4.15 [16, Lemma 3.1] *Let $\alpha \in \left]0,+\infty\right[$, let $\mu \in \left]0,+\infty\right[$, and let A and B be operators in $\mathcal{S}(\mathcal{H})$ such that $\mu\,\mathrm{Id} \succcurlyeq A \succcurlyeq B \succcurlyeq \alpha\,\mathrm{Id}$. Then the following hold :*

(i) $\alpha^{-1}\,\mathrm{Id} \succcurlyeq B^{-1} \succcurlyeq A^{-1} \succcurlyeq \mu^{-1}\,\mathrm{Id}$.

(ii) $(\forall x \in \mathcal{H})\ \langle A^{-1}x \mid x\rangle \geqslant \|A\|^{-1}\|x\|^2$.

(iii) $\|A^{-1}\| \leqslant \alpha^{-1}$.

4.2.3 Main convergence result

We present our main result.

Theorem 4.16 *Let $\alpha \in \left]0,+\infty\right[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that*

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \textit{and} \quad (\forall n \in \mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.28}$$

Let $\varepsilon \in \left]0,1\right[$, let $m \geqslant 1$ be an integer, and let $x_0 \in \mathcal{H}$. For every $i \in \{1,\ldots,m\}$ and every $n \in \mathbb{N}$, let $\alpha_{i,n} \in \left]0,1\right[$, let $T_{i,n}\colon \mathcal{H} \to \mathcal{H}$ be $\alpha_{i,n}$-averaged on $(\mathcal{H}, U_n^{-1})$, let ϕ_n be an averagedness constant of $T_{1,n}\cdots T_{m,n}$, let $\lambda_n \in \left]0,\phi_n\right]$, and let $e_{i,n} \in \mathcal{H}$. Iterate

$$\begin{array}{l}\text{for } n = 0,1,\ldots\\ \left\lfloor\begin{array}{l} y_n = T_{1,n}\Big(T_{2,n}\big(\cdots T_{m-1,n}(T_{m,n}x_n + e_{m,n}) + e_{m-1,n}\cdots\big) + e_{2,n}\Big) + e_{1,n}\\ x_{n+1} = x_n + \lambda_n(y_n - x_n).\end{array}\right.\end{array} \tag{4.29}$$

Suppose that

$$S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix}(T_{1,n}\cdots T_{m,n}) \neq \varnothing \tag{4.30}$$

and

$$(\forall i \in \{1,\ldots,m\}) \quad \sum_{n\in\mathbb{N}} \lambda_n \|e_{i,n}\|_{U_n^{-1}} < +\infty, \tag{4.31}$$

and define

$$(\forall i \in \{1,\ldots,m\})(\forall n \in \mathbb{N}) \quad T_{i+,n} = \begin{cases} T_{i+1,n}\cdots T_{m,n}, & \text{if } i \neq m; \\ \mathrm{Id}\,, & \text{if } i = m. \end{cases} \tag{4.32}$$

Then the following hold :

(i) $\sum_{n\in\mathbb{N}} \lambda_n(1/\phi_n - \lambda_n)\|T_{1,n}\cdots T_{m,n}x_n - x_n\|^2_{U_n^{-1}} < +\infty$.

(ii) *Suppose that* $(\forall n \in \mathbb{N})\ \lambda_n \in \left]0, \varepsilon + (1-\varepsilon)/\phi_n\right]$. *Then* $(\forall x \in S)$

$$\max_{1\leqslant i\leqslant m} \sum_{n\in\mathbb{N}} \frac{\lambda_n(1-\alpha_{i,n})}{\alpha_{i,n}} \|(\mathrm{Id}\,-T_{i,n})T_{i+,n}x_n - (\mathrm{Id}\,-T_{i,n})T_{i+,n}x\|^2_{U_n^{-1}} < +\infty. \tag{4.33}$$

(iii) $(x_n)_{n\in\mathbb{N}}$ *converges weakly to a point in* S *if and only if every weak sequential cluster point of* $(x_n)_{n\in\mathbb{N}}$ *is in* S. *In this case, the convergence is strong if* $\operatorname{int} S \neq \varnothing$.

(iv) $(x_n)_{n\in\mathbb{N}}$ *converges strongly to a point in* S *if and only if* $\underline{\lim}\, d_S(x_n) = 0$.

Proof. Let $n \in \mathbb{N}$ and let $x \in S$. Set

$$T_n = T_{1,n}\cdots T_{m,n} \tag{4.34}$$

and

$$e_n = y_n - T_n x_n. \tag{4.35}$$

Using the nonexpansiveness on $(\mathcal{H}, U_n^{-1})$ of the operators $(T_{i,n})_{1\leqslant i\leqslant m}$, we first derive from (4.35) that

$$\|e_n\|_{U_n^{-1}} \leqslant \sum_{i=1}^m \|e_{i,n}\|_{U_n^{-1}}. \tag{4.36}$$

Let us rewrite (4.29) as

$$x_{n+1} = x_n + \lambda_n\big(T_n x_n + e_n - x_n\big), \tag{4.37}$$

and set

$$R_n = (1 - 1/\phi_n)\,\mathrm{Id} + (1/\phi_n)T_n \quad \text{and} \quad \mu_n = \phi_n\lambda_n. \tag{4.38}$$

Then $\operatorname{Fix} R_n = \operatorname{Fix} T_n$ and, by Lemmas 4.8 and 4.9, R_n is nonexpansive on $(\mathcal{H}, U_n^{-1})$. Furthermore, (4.37) can be written as

$$x_{n+1} = x_n + \mu_n\big(R_n x_n - x_n\big) + \lambda_n e_n, \quad \text{where} \quad \mu_n \in \left]0,1\right[. \tag{4.39}$$

Now set $z_n = x_n + \mu_n(R_n x_n - x_n)$. Since $x \in \operatorname{Fix} R_n$, we derive from [4, Corollary 2.14] that

$$\begin{aligned}\|z_n - x\|^2_{U_n^{-1}} &= (1-\mu_n)\|x_n - x\|^2_{U_n^{-1}} + \mu_n\|R_n x_n - R_n x\|^2_{U_n^{-1}} \\ &\quad - \mu_n(1-\mu_n)\|R_n x_n - x_n\|^2_{U_n^{-1}} &(4.40)\\ &\leqslant \|x_n - x\|^2_{U_n^{-1}} - \lambda_n(1/\phi_n - \lambda_n)\|T_n x_n - x_n\|^2_{U_n^{-1}}. &(4.41)\end{aligned}$$

Hence, (4.39), (4.28), and (4.41) yield

$$\begin{aligned}\|x_{n+1} - x\|_{U_{n+1}^{-1}} &\leqslant \sqrt{1+\eta_n}\|z_n - x\|_{U_n^{-1}} + \lambda_n\sqrt{1+\eta_n}\|e_n\|_{U_n^{-1}} &(4.42)\\ &\leqslant \sqrt{1+\eta_n}\|x_n - x\|_{U_n^{-1}} + \lambda_n\sqrt{1+\eta_n}\|e_n\|_{U_n^{-1}} &(4.43)\end{aligned}$$

and, since $\sum_{k\in\mathbb{N}} \lambda_k\|e_k\|_{U_k} < +\infty$, it follows from Lemma 4.11 that

$$\nu = \sum_{k\in\mathbb{N}} \lambda_k\|e_k\|_{U_k^{-1}} + 2\sup_{k\in\mathbb{N}}\|x_k - x\|_{U_k^{-1}} < +\infty. \tag{4.44}$$

Moreover, using (4.42) and (4.41) we write

$$\begin{aligned}(1+\eta_n)^{-1}\|x_{n+1} - x\|^2_{U_{n+1}^{-1}} &\leqslant \|z_n - x\|^2_{U_n^{-1}} + (2\|z_n - x\|_{U_n^{-1}} + \lambda_n\|e_n\|_{U_n^{-1}})\lambda_n\|e_n\|_{U_n^{-1}} &(4.45)\\ &\leqslant \|x_n - x\|^2_{U_n^{-1}} - \lambda_n(1/\phi_n - \lambda_n)\|T_n x_n - x_n\|^2_{U_n^{-1}} \\ &\quad + \nu\lambda_n\|e_n\|_{U_n^{-1}}. &(4.46)\end{aligned}$$

(i) : This follows from (4.46), (4.34), (4.30), (4.44), and Lemma 4.10.

(ii) : We apply the definition of averagedness of the operators $(T_{i,n})_{1\leqslant i\leqslant m}$ to obtain

$$\begin{aligned}\|T_n x_n - x\|^2_{U_n^{-1}} &= \|T_{1,n}\cdots T_{m,n}x_n - T_{1,n}\cdots T_{m,n}x\|^2_{U_n^{-1}} \\ &\leqslant \|T_{2,n}\cdots T_{m,n}x_n - T_{2,n}\cdots T_{m,n}x\|^2_{U_n^{-1}} \\ &\quad - \frac{1-\alpha_{1,n}}{\alpha_{1,n}}\|(\mathrm{Id} - T_{1,n})T_{2,n}\cdots T_{m,n}x_n - (\mathrm{Id} - T_{1,n})T_{2,n}\cdots T_{m,n}x\|^2_{U_n^{-1}} \\ &\;\;\vdots \\ &\leqslant \|x_n - x\|^2_{U_n^{-1}} \\ &\quad - \sum_{i=1}^m \frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id} - T_{i,n})T_{i+,n}x_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2_{U_n^{-1}}. &(4.47)\end{aligned}$$

Note also that

$$\lambda_n \leqslant \varepsilon + \frac{1-\varepsilon}{\phi_n} \;\Rightarrow\; \frac{1}{\varepsilon}\lambda_n \leqslant (\frac{1}{\varepsilon} - 1)\frac{1}{\phi_n}$$
$$\Leftrightarrow\; \lambda_n - 1 \leqslant \left(\frac{1}{\varepsilon} - 1\right)\left(\frac{1}{\phi_n} - \lambda_n\right). \tag{4.48}$$

Thus (4.45), the definition of z_n, and [4, Corollary 2.14] yield

$$\begin{aligned}(1+\eta_n)^{-1}\|x_{n+1} - x\|^2_{U^{-1}_{n+1}} &\leqslant \|(1-\lambda_n)(x_n - x) + \lambda_n(T_n x_n - x)\|^2_{U_n^{-1}} + \nu\lambda_n\|e_n\|_{U_n^{-1}} \\ &= (1-\lambda_n)\|x_n - x\|^2_{U_n^{-1}} + \lambda_n\|T_n x_n - x\|^2_{U_n^{-1}} \\ &\quad + \lambda_n(\lambda_n - 1)\|T_n x_n - x_n\|^2_{U_n^{-1}} + \nu\lambda_n\|e_n\|_{U_n^{-1}} \\ &\leqslant (1-\lambda_n)\|x_n - x\|^2_{U_n^{-1}} + \lambda_n\|T_n x_n - x\|^2_{U_n^{-1}} + \varepsilon_n, \end{aligned} \tag{4.49}$$

where

$$\varepsilon_n = \lambda_n\left(\frac{1}{\varepsilon} - 1\right)\left(\frac{1}{\alpha_n} - \lambda_n\right)\|T_n x_n - x_n\|^2_{U_n^{-1}} + \nu\lambda_n\|e_n\|_{U_n^{-1}}. \tag{4.50}$$

Now set

$$\beta_n = \lambda_n \max_{1\leqslant i\leqslant m}\left(\frac{1-\alpha_{i,n}}{\alpha_{i,n}}\|(\mathrm{Id} - T_{i,n})T_{i+,n}x_n - (\mathrm{Id} - T_{i,n})T_{i+,n}x\|^2_{U_n^{-1}}\right). \tag{4.51}$$

On the one hand, it follows from (i), (4.44), and (4.30) that

$$\sum_{k\in\mathbb{N}} \varepsilon_k < +\infty. \tag{4.52}$$

On the other hand, combining (4.47) and (4.49), we obtain

$$(1+\eta_n)^{-1}\|x_{n+1} - x\|^2_{U^{-1}_{n+1}} \leqslant \|x_n - x\|^2_{U_n^{-1}} - \beta_n + \varepsilon_n. \tag{4.53}$$

Consequently, Lemma 4.10 implies that $\sum_{k\in\mathbb{N}}\beta_k < +\infty$.

(iii)–(iv) : The results follow from (4.53), (4.52), and Proposition 4.12 for the weak convergence, and Propositions 4.13 and 4.14 for the strong convergence. □

Remark 4.17 Suppose that $(\forall n \in \mathbb{N})\ U_n = \mathrm{Id}$ and $\lambda_n \leqslant (1-\varepsilon)(1/\phi_n + \varepsilon)$. Then Theorem 4.16 reduces to [18, Theorem 3.5] which itself extends [9, Section 3] in the case $(\forall n \in \mathbb{N})\ \lambda_n \leqslant 1$. As far as we know, it is the first inexact overrelaxed variable metric algorithm based on averaged operators.

4.2.4 Applications to the forward-backward algorithm

A special case of Theorem 4.16 of interest is the following.

Corollary 4.18 *Let $\alpha \in \left]0,+\infty\right[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that*

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \textit{and} \quad (\forall n\in\mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.54}$$

Let $\varepsilon \in \left]0,1\right[$ and let $x_0 \in \mathcal{H}$. For every $n \in \mathbb{N}$, let $\alpha_{1,n} \in \left]0,1/(1+\varepsilon)\right]$, let $\alpha_{2,n} \in \left]0,1/(1+\varepsilon)\right]$, let $T_{1,n}\colon \mathcal{H} \to \mathcal{H}$ be $\alpha_{1,n}$-averaged on $(\mathcal{H}, U_n^{-1})$, let $T_{2,n}\colon \mathcal{H} \to \mathcal{H}$ be $\alpha_{2,n}$-averaged on $(\mathcal{H}, U_n^{-1})$, let $e_{1,n} \in \mathcal{H}$, and let $e_{2,n} \in \mathcal{H}$. In addition, for every $n \in \mathbb{N}$, let

$$\lambda_n \in \left[\varepsilon, \varepsilon + \frac{1-\varepsilon}{\phi_n}\right], \quad \textit{where} \quad \phi_n = \frac{\alpha_{1,n}+\alpha_{2,n}-2\alpha_{1,n}\alpha_{2,n}}{1-\alpha_{1,n}\alpha_{2,n}}, \tag{4.55}$$

and iterate

$$x_{n+1} = x_n + \lambda_n\Big(T_{1,n}\big(T_{2,n}x_n + e_{2,n}\big) + e_{1,n} - x_n\Big). \tag{4.56}$$

Suppose that

$$S = \bigcap_{n\in\mathbb{N}} \operatorname{Fix}(T_{1,n}T_{2,n}) \neq \varnothing, \quad \sum_{n\in\mathbb{N}} \lambda_n\|e_{1,n}\| < +\infty, \quad \textit{and} \quad \sum_{n\in\mathbb{N}} \lambda_n\|e_{2,n}\| < +\infty. \tag{4.57}$$

Then the following hold :

(i) $\sum_{n\in\mathbb{N}} \|T_{1,n}T_{2,n}x_n - x_n\|^2 < +\infty$.

(ii) $(\forall x \in S)$ $\sum_{n\in\mathbb{N}} \|T_{1,n}T_{2,n}x_n - T_{2,n}x_n + T_{2,n}x - x\|^2 < +\infty$.

(iii) $(\forall x \in S)$ $\sum_{n\in\mathbb{N}} \|T_{2,n}x_n - x_n - T_{2,n}x + x\|^2 < +\infty$.

(iv) *Suppose that every weak sequential cluster point of $(x_n)_{n\in\mathbb{N}}$ is in S. Then $(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in S, and the convergence is strong if $\operatorname{int} S \neq \varnothing$.*

(v) *$(x_n)_{n\in\mathbb{N}}$ converges strongly to a point in S if and only if $\varliminf d_S(x_n) = 0$.*

Proof. For every $n \in \mathbb{N}$,

$$\frac{1}{\sqrt{\mu}}\|e_{1,n}\| \leqslant \|e_{1,n}\|_{U_n^{-1}} \quad \text{and} \quad \frac{1}{\sqrt{\mu}}\|e_{2,n}\| \leqslant \|e_{2,n}\|_{U_n^{-1}}, \tag{4.58}$$

and $T_{1,n}T_{2,n}$ is ϕ_n-averaged by [4, Proposition 4.44]. Thus, we apply Theorem 4.16 with $m = 2$.

(i)–(iii) : This follows from Theorem 4.16(i) that

$$(\forall x \in S) \quad \begin{cases} \sum_{n\in\mathbb{N}} \dfrac{\lambda_n(1-\alpha_{1,n})}{\alpha_{1,n}} \|(\mathrm{Id}\, - T_{1,n})T_{2,n}x_n - (\mathrm{Id}\, - T_{1,n})T_{2,n}x\|^2_{U_n^{-1}} < +\infty \\ \sum_{n\in\mathbb{N}} \dfrac{\lambda_n(1-\alpha_{2,n})}{\alpha_{2,n}} \|(\mathrm{Id}\, - T_{2,n})x_n - (\mathrm{Id}\, - T_{2,n})x\|^2_{U_n^{-1}} < +\infty \\ \sum_{n\in\mathbb{N}} \lambda_n\Big(\dfrac{1}{\phi_n} - \lambda_n\Big) \|T_{1,n}T_{2,n}x_n - x_n\|^2_{U_n^{-1}} < +\infty. \end{cases} \tag{4.59}$$

However, we derive from the assumptions that

$$(\forall x \in S)(\forall n \in \mathbb{N}) \quad \begin{cases} T_{1,n}T_{2,n}x = x \\ \dfrac{\lambda_n(1-\alpha_{1,n})}{\alpha_{1,n}} \geqslant \varepsilon^2 \\ \dfrac{\lambda_n(1-\alpha_{2,n})}{\alpha_{2,n}} \geqslant \varepsilon^2 \\ \lambda_n\Big(\dfrac{1}{\phi_n} - \lambda_n\Big) \geqslant \varepsilon\dfrac{1-\phi_n}{\phi_n} \geqslant \dfrac{2\varepsilon^2}{2\varepsilon+1}. \end{cases} \tag{4.60}$$

Combining (4.54), (4.59) and (4.60) completes the proof.

(iv)–(v) : This follows from Theorem 4.16(iii)–(iv). □

Remark 4.19 This corollary is a variable metric version of [18, Corollary 4.1] where $(\forall n \in \mathbb{N})\ U_n = \mathrm{Id}$ and $\lambda_n \leqslant (1-\varepsilon)(1/\phi_n + \varepsilon)$.

We recall the definition of a demiregular operator. See [2] for examples of demiregular operators.

Definition 4.20 [2, Definition 2.3] An operator $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ is *demiregular* at $x \in \mathrm{dom}\, A$ if, for every sequence $((x_n, u_n))_{n\in\mathbb{N}}$ in $\mathrm{gra}\, A$ and every $u \in Ax$ such that $x_n \rightharpoonup x$ and $u_n \to u$, we have $x_n \to x$.

Proposition 4.21 *Let $\alpha \in\,]0,+\infty[$, let $U \in \mathcal{P}_\alpha(\mathcal{H})$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\beta \in\,]0,+\infty[$, let $\gamma \in\,]0, 2\beta/\|U\|]$, and let B a β-cocoercive operator. Then the following hold :*

(i) *$J_{\gamma UA}$ is a $1/2$-averaged operator on $(\mathcal{H}, U^{-1})$.*

(ii) *$\mathrm{Id}\, - \gamma UB$ is a $\gamma\|U\|/(2\beta)$-averaged operator on $(\mathcal{H}, U^{-1})$.*

Proof. (i) : [16, Lemma 3.7].

(ii) : We derive from (4.21) and Lemma 4.15(iii) that for every $x \in \mathcal{H}$ and for every $y \in \mathcal{H}$

$$\begin{aligned} \langle x - y \mid UBx - UBy \rangle_{U^{-1}} &= \langle x - y \mid Bx - By \rangle \\ &\geqslant \beta \langle Bx - By \mid Bx - By \rangle \\ &= \beta \big\langle U^{-1}(UBx - UBy) \mid UBx - UBy \big\rangle_{U^{-1}} \\ &\geqslant \|U\|^{-1} \beta \|UBx - UBy\|^2_{U^{-1}}. \end{aligned} \tag{4.61}$$

Thus, for every $x \in \mathcal{H}$ and for every $y \in \mathcal{H}$

$$\begin{aligned} \|(x - \gamma UBx) - (y - \gamma UBy)\|^2_{U^{-1}} &= \|x - y\|^2_{U^{-1}} + \|\gamma UBx - \gamma UBy\|^2_{U^{-1}} \\ &\qquad - 2\gamma \langle x - y \mid UBx - UBy \rangle_{U^{-1}} \qquad (4.62) \\ &\leqslant \|x - y\|^2_{U^{-1}} \\ &\qquad - \gamma(2\beta/\|U\| - \gamma)\|UBx - UBy\|^2_{U^{-1}}, \qquad (4.63) \end{aligned}$$

which concludes the proof. ☐

Next, we introduce a new variable metric forward-backward splitting algorithm.

Proposition 4.22 *Let $\beta \in \left]0, +\infty\right[$, let $\varepsilon \in \left]0, \min\{1/2, \beta\}\right[$, let $\alpha \in \left]0, +\infty\right[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell^1_+(\mathbb{N})$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that*

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad (1 + \eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.64}$$

Let $x_0 \in \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, and let $B\colon \mathcal{H} \to \mathcal{H}$ be β-cocoercive. Furthermore, let $(a_n)_{n\in\mathbb{N}}$ and $(b_n)_{n\in\mathbb{N}}$ be sequences in $\mathcal{H}$ such that $\sum_{n\in\mathbb{N}} \|a_n\| < +\infty$ and $\sum_{n\in\mathbb{N}} \|b_n\| < +\infty$. Suppose that $\operatorname{zer}(A + B) \neq \varnothing$ and, for every $n \in \mathbb{N}$, let

$$\gamma_n \in \left[\varepsilon, \frac{2\beta}{(1+\varepsilon)\|U_n\|}\right] \quad \text{and} \quad \lambda_n \in \left[\varepsilon, 1 + (1 - \varepsilon)\left(1 - \frac{\gamma_n\|U_n\|}{2\beta}\right)\right], \tag{4.65}$$

and iterate

$$x_{n+1} = x_n + \lambda_n\Big(J_{\gamma_n U_n A}\big(x_n - \gamma_n U_n(Bx_n + b_n)\big) + a_n - x_n\Big). \tag{4.66}$$

Then the following hold :

(i) $\sum_{n\in\mathbb{N}} \|J_{\gamma_n U_n A}(x_n - \gamma_n U_n Bx_n) - x_n\|^2 < +\infty$.

(ii) *Let $x \in \operatorname{zer}(A + B)$. Then $\sum_{n\in\mathbb{N}} \|Bx_n - Bx\|^2 < +\infty$.*

(iii) *$(x_n)_{n\in\mathbb{N}}$ converges weakly to a point in $\operatorname{zer}(A + B)$.*

(iv) *Suppose that one of the following holds :*

  (a) *A is demiregular at every point in $\operatorname{zer}(A + B)$.*

  (b) *B is demiregular at every point in $\operatorname{zer}(A + B)$.*

(c) $\operatorname{int} S \neq \varnothing$.

Then $(x_n)_{n\in\mathbb{N}}$ *converges strongly to a point in* $\operatorname{zer}(A+B)$.

Proof. We apply Corollary 4.18. Set

$$(\forall n \in \mathbb{N}) \;\; T_{1,n} = J_{\gamma_n U_n A}, \quad T_{2,n} = \operatorname{Id} - \gamma_n U_n B, \quad e_{1,n} = a_n, \quad \text{and} \quad e_{2,n} = -\gamma_n U_n b_n. \tag{4.67}$$

Then, for every $n \in \mathbb{N}$, $T_{1,n}$ is $\alpha_{1,n}$-averaged on $(\mathcal{H}, U_n^{-1})$ with $\alpha_{1,n} = 1/2$ and $T_{2,n}$ is $\alpha_{2,n}$-averaged on $(\mathcal{H}, U_n^{-1})$ with $\alpha_{2,n} = \gamma_n\|U_n\|/(2\beta)$ by Proposition 4.21. Moreover, for every $n \in \mathbb{N}$,

$$\phi_n = \frac{\alpha_{1,n} + \alpha_{2,n} - 2\alpha_{1,n}\alpha_{2,n}}{1 - \alpha_{1,n}\alpha_{2,n}} = \frac{2\beta}{4\beta - \gamma_n\|U_n\|} \tag{4.68}$$

and, therefore, (4.65) yields

$$\lambda_n \in \left[\varepsilon, \varepsilon + \frac{1-\varepsilon}{\phi_n}\right]. \tag{4.69}$$

Hence, we derive from (4.68) and (4.69) that $(\forall n \in \mathbb{N})$ $\lambda_n \leqslant 2 + \varepsilon$. Consequently,

$$\begin{cases} \sum_{n\in\mathbb{N}} \lambda_n\|e_{1,n}\| = (2+\varepsilon)\sum_{n\in\mathbb{N}}\|a_n\| < +\infty \\ \sum_{n\in\mathbb{N}} \lambda_n\|e_{2,n}\| \leqslant 2\beta(2+\varepsilon)\mu\alpha^{-1}\sum_{n\in\mathbb{N}}\|b_n\| < +\infty. \end{cases} \tag{4.70}$$

Furthermore, it follows from [4, Proposition 26.1(iv)] that

$$(\forall n \in \mathbb{N}) \quad S = \operatorname{zer}(A+B) = \operatorname{Fix}(T_{1,n}T_{2,n}) \neq \varnothing. \tag{4.71}$$

Hence, the assumptions of Corollary 4.18 are satisfied.

(i) : This is a consequence of Corollary 4.18(i) and (4.67).

(ii) : Corollary 4.18(ii), (4.67), and Lemma 4.15(iii) yield

$$\begin{aligned} \sum_{n\in\mathbb{N}} \|Bx_n - Bx\|^2 &= \sum_{n\in\mathbb{N}} \gamma_n^{-2}\|U_n^{-1}(T_{2,n}x_n - x_n - T_{2,n}x + x)\|^2 \\ &\leqslant \frac{1}{\varepsilon^2\alpha^2}\sum_{n\in\mathbb{N}}\|T_{2,n}x_n - x_n - T_{2,n}x + x\|^2 \\ &< +\infty. \end{aligned} \tag{4.72}$$

(iii) : Let $(k_n)_{n\in\mathbb{N}}$ be a strictly increasing sequence in $\mathbb{N}$ and let $y \in \mathcal{H}$ be such that $x_{k_n} \rightharpoonup y$. In view of Corollary 4.18(iv), it remains to show that $y \in \operatorname{zer}(A+B)$. Set

$$(\forall n \in \mathbb{N}) \quad \begin{cases} y_n = J_{\gamma_n U_n A}(x_n - \gamma_n U_n B x_n) \\ u_n = \gamma_n^{-1}U_n^{-1}(x_n - y_n) - Bx_n \\ v_n = Bx_n \end{cases} \tag{4.73}$$

and let $z \in \operatorname{zer}(A+B)$. Hence, we derive from (i) that $y_n - x_n \to 0$. Then $y_{k_n} \rightharpoonup y$ and by (ii) $Bx_n \to Bz$. Altogether, $y_{k_n} \rightharpoonup y$, $v_{k_n} \rightharpoonup Bz$, $y_{k_n} - x_{k_n} \to 0$, $u_{k_n} + v_{k_n} \to 0$, and, for every $n \in \mathbb{N}$, $u_{k_n} \in Ay_{k_n}$ and $v_{k_n} \in Bx_{k_n}$. It therefore follows from [11, Lemma 4.5(ii)] that $y \in \operatorname{zer}(A+B)$.

(iv) : The proof is the same that in [18, Proposition 4.4(iv)]. □

Remark 4.23 Suppose that $(\forall n \in \mathbb{N})\ U_n = \operatorname{Id}$ and $\lambda_n \leqslant (1-\varepsilon)(1/\phi_n + \varepsilon)$. Then Proposition 4.22 captures [18, Proposition 4.4]. Now suppose that $(\forall n \in \mathbb{N})\ \lambda_n \leqslant 1$. Then Proposition 4.22 captures [16, Theorem 4.1].

Using the averaged operators framework allows us to obtain an extended forward-backward splitting algorithm in Euclidean spaces.

Example 4.24 Let $\alpha \in]0, +\infty[$, let $(\eta_n)_{n\in\mathbb{N}} \in \ell_+^1(\mathbb{N})$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that

$$\mu = \sup_{n\in\mathbb{N}} \|U_n\| < +\infty \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad (1+\eta_n)U_{n+1} \succcurlyeq U_n. \tag{4.74}$$

Let $\varepsilon \in]0, 1/2[$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be a maximally monotone operator, let $\beta \in]0, +\infty[$, let B a β-cocoercive operator, and let $(\gamma_n)_{n\in\mathbb{N}}$ and $(\mu_n)_{n\in\mathbb{N}}$ be sequences in $[\varepsilon, +\infty[$ such that

$$\phi_n = \frac{2\mu_n\beta}{4\beta - \|U_n\|\gamma_n} \leqslant 1 - \varepsilon. \tag{4.75}$$

Let $x_0 \in \mathcal{H}$ and iterate

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \mu_n\Big(J_{\gamma_n U_n A}(x_n - \gamma_n U_n B) - x_n\Big). \tag{4.76}$$

Suppose that $\mathcal{H}$ is finite-dimensional and that $\operatorname{zer}(A+B) \neq \varnothing$. Then $(x_n)_{n\in\mathbb{N}}$ converges to a point in $\operatorname{zer}(A+B)$.

Proof. Set $(\forall n \in \mathbb{N})\ T_n = \operatorname{Id} + \mu_n(J_{\gamma_n U_n A}(\operatorname{Id} - \gamma_n U_n B) - \operatorname{Id})$. Remark that, for every $n \in \mathbb{N}$, T_n is ϕ_n-averaged. Hence we apply Theorem 4.16 with $m = 1$ and $\lambda \equiv 1$. It follows from Theorem 4.16(i) and (4.75) that $T_n x_n - x_n \to 0$. Since $\mathcal{H}$ is finite-dimensional, the claim follows from Theorem 4.16(iii). □

Remark 4.25 An underrelaxation or an appropriate choice of the metric of the algorithm allows us to exceed the classical bound $2/\beta$ for $(\gamma_n)_{n\in\mathbb{N}}$. For instance, the parameters $\gamma_n \equiv 2.99/\beta$, $\mu_n \equiv 1/2$, and $U_n \equiv \operatorname{Id}$ satisfy the assumptions.

4.2.5 A composite monotone inclusion problem

We study the composite monotone inclusion presented in [14].

Problem 4.26 Let $z \in \mathcal{H}$, let $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ be maximally monotone, let $\mu \in \,]0,+\infty[$, let $C\colon \mathcal{H} \to \mathcal{H}$ be μ-cocoercive, and let m be a strictly positive integer. For every $i \in \{1,\ldots,m\}$, let $r_i \in \mathcal{G}_i$, let $B_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone, let $\nu_i \in \,]0,+\infty[$, let $D_i\colon \mathcal{G}_i \to 2^{\mathcal{G}_i}$ be maximally monotone and ν_i-strongly monotone, and suppose that $0 \neq L_i \in \mathcal{B}(\mathcal{H},\mathcal{G}_i)$. The problem is to find $\overline{x} \in \mathcal{H}$ such that

$$z \in A\overline{x} + \sum_{i=1}^{m} L_i^*\big((B_i \,\Box\, D_i)(L_i\overline{x} - r_i)\big) + C\overline{x}, \tag{4.77}$$

the dual problem of which is to find $\overline{v}_1 \in \mathcal{G}_1, \ldots, \overline{v}_m \in \mathcal{G}_m$ such that

$$(\exists\, x \in \mathcal{H}) \quad \begin{cases} z - \sum_{i=1}^{m} L_i^*\overline{v}_i \in Ax + Cx \\ (\forall i \in \{1,\ldots,m\})\ \overline{v}_i \in (B_i \,\Box\, D_i)(L_i x - r_i). \end{cases} \tag{4.78}$$

The following corollary is an overrelaxed version of [16, Corollary 6.2].

Corollary 4.27 *In Problem 4.26, suppose that*

$$z \in \operatorname{ran}\left(A + \sum_{i=1}^{m} L_i^*\big((B_i \,\Box\, D_i)(L_i \cdot - r_i)\big) + C\right), \tag{4.79}$$

and set

$$\beta = \min\{\mu, \nu_1, \ldots, \nu_m\}. \tag{4.80}$$

Let $\varepsilon \in \,]0, \min\{1,\beta\}[$, let $\alpha \in \,]0,+\infty[$, let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$, let $x_0 \in \mathcal{H}$, let $(a_n)_{n\in\mathbb{N}}$ and $(c_n)_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$, and let $(U_n)_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{H})$ such that $(\forall n \in \mathbb{N})$ $U_{n+1} \succcurlyeq U_n$. For every $i \in \{1,\ldots,m\}$, let $v_{i,0} \in \mathcal{G}_i$, and let $(b_{i,n})_{n\in\mathbb{N}}$ and $(d_{i,n})_{n\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{G}_i$, and let $(U_{i,n})_{n\in\mathbb{N}}$ be a sequence in $\mathcal{P}_\alpha(\mathcal{G}_i)$ such that $(\forall n \in \mathbb{N})$ $U_{i,n+1} \succcurlyeq U_{i,n}$. For every $n \in \mathbb{N}$, set

$$\delta_n = \left(\sqrt{\sum_{i=1}^{m} \|\sqrt{U_{i,n}}L_i\sqrt{U_n}\|^2}\right)^{-1} - 1, \tag{4.81}$$

suppose that

$$\zeta_n = \frac{1+\delta_n}{(1+\delta_n)\max\{\|U_n\|, \|U_{1,n}\|, \ldots, \|U_{m,n}\|\}} \geqslant \frac{1}{2\beta - \varepsilon}, \tag{4.82}$$

and let

$$\lambda_n \in \left[\varepsilon, 1 + (1-\varepsilon)\left(1 - \frac{1}{2\zeta_n\beta}\right)\right]. \tag{4.83}$$

Iterate

$$
\begin{array}{l}
\textit{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
p_n = J_{U_n A}\Big(x_n - U_n\big(\sum_{i=1}^m L_i^* v_{i,n} + Cx_n + c_n - z\big)\Big) + a_n \\
y_n = 2p_n - x_n \\
x_{n+1} = x_n + \lambda_n(p_n - x_n) \\
\text{for } i = 1, \ldots, m \\
\left\lfloor
\begin{array}{l}
q_{i,n} = J_{U_{i,n} B_i^{-1}}\Big(v_{i,n} + U_{i,n}\big(L_i y_n - D_i^{-1} v_{i,n} - d_{i,n} - r_i\big)\Big) + b_{i,n} \\
v_{i,n+1} = v_{i,n} + \lambda_n(q_{i,n} - v_{i,n}).
\end{array}
\right.
\end{array}
\right.
\end{array}
\tag{4.84}
$$

Then the following hold for some solution $\overline{x}$ *to* (4.77) *and some solution* $(\overline{v}_1, \ldots, \overline{v}_m)$ *to* (4.78) *:*

(i) $x_n \rightharpoonup \overline{x}$.

(ii) $(\forall i \in \{1, \ldots, m\})$ $v_{i,n} \rightharpoonup \overline{v}_i$.

(iii) *Suppose that* C *is demiregular at* $\overline{x}$*. Then* $x_n \to \overline{x}$.

(iv) *Suppose that, for some* $j \in \{1, \ldots, m\}$*,* D_j^{-1} *is demiregular at* $\overline{v}_j$*. Then* $v_{j,n} \to \overline{v}_j$.

Proof. Set $\boldsymbol{\mathcal{G}} = \mathcal{G}_1 \oplus \cdots \oplus \mathcal{G}_m$, $\boldsymbol{\mathcal{K}} = \mathcal{H} \oplus \boldsymbol{\mathcal{G}}$, and

$$
\begin{cases}
\widetilde{A}\ : \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}} \colon (x, v_1, \ldots, v_m) \mapsto \big(\sum_{i=1}^m L_i^* v_i - z + Ax\big) \\
\qquad\qquad\qquad\qquad \times (r_1 - L_1 x + B_1^{-1} v_1) \times \cdots \times (r_m - L_m x + B_m^{-1} v_m) \\
\widetilde{B}\ : \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon (x, v_1, \ldots, v_m) \mapsto \big(Cx, D_1^{-1} v_1, \ldots, D_m^{-1} v_m\big) \\
\widetilde{S}\ : \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon (x, v_1, \ldots, v_m) \mapsto \Big(\sum_{i=1}^m L_i^* v_i, -L_1 x, \ldots, -L_m x\Big).
\end{cases}
\tag{4.85}
$$

Now, for every $n \in \mathbb{N}$, define

$$
\begin{cases}
\widetilde{U}_n \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon (x, v_1, \ldots, v_m) \mapsto \Big(U_n x, U_{1,n} v_1, \ldots, U_{m,n} v_m\Big) \\
\widetilde{V}_n \colon \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}} \colon \\
\qquad (x, v_1, \ldots, v_m) \mapsto \Big(U_n^{-1} x - \sum_{i=1}^m L_i^* v_i, \big(-L_i x + U_{i,n}^{-1} v_i\big)_{1 \leqslant i \leqslant m}\Big)
\end{cases}
\tag{4.86}
$$

and

$$
\begin{cases}
\widetilde{x}_n = (x_n, v_{1,n}, \ldots, v_{m,n}) \\
\widetilde{y}_n = (p_n, q_{1,n}, \ldots, q_{m,n}) \\
\widetilde{a}_n = (a_n, b_{1,n}, \ldots, b_{m,n}) \\
\widetilde{c}_n = (c_n, d_{1,n}, \ldots, d_{m,n}) \\
\widetilde{d}_n = (U_n^{-1} a_n, U_{1,n}^{-1} b_{1,n}, \ldots, U_{m,n}^{-1} b_{m,n})
\end{cases}
\quad \text{and} \quad \widetilde{b}_n = (\widetilde{S} + \widetilde{V}_n)\widetilde{a}_n + \widetilde{c}_n - \widetilde{d}_n.
\tag{4.87}
$$

It follows from the proof of [16, Corollary 6.2] that (4.84) is equivalent to

$$(\forall n \in \mathbb{N}) \quad \widetilde{x}_{n+1} = \widetilde{x}_n + \lambda_n \Big(J_{\widetilde{V}_n^{-1}\widetilde{A}}\big(\widetilde{x}_n - \widetilde{V}_n^{-1}(\widetilde{B}\widetilde{x}_n + \widetilde{b}_n)\big) + \widetilde{a}_n - \widetilde{x}_n \Big), \tag{4.88}$$

that the operators $\widetilde{A}$ and $\widetilde{B}$ are maximally monotone, and $\widetilde{B}$ is β-cocoercive on $\boldsymbol{\mathcal{H}}$. Furthermore, for every $(\overline{x}, \overline{v}) \in \operatorname{zer}(\widetilde{A} + \widetilde{B})$, $\overline{x}$ solves (4.77) and $\overline{v}$ solves (4.78). Now set $\rho = 1/\alpha + \sqrt{\sum_{i=1}^m \|L_i\|^2}$. We deduce from the proof of [16, Corollary 6.2] that $(\forall n \in \mathbb{N})$ $\|\widetilde{V}_n^{-1}\| \leqslant \zeta_n^{-1} \leqslant 2\beta - \varepsilon$ and $\widetilde{V}_{n+1}^{-1} \succcurlyeq \widetilde{V}_n^{-1} \in \mathcal{P}_{1/\rho}(\boldsymbol{\mathcal{K}})$. We observe that (4.88) has the structure of the variable metric forward-backward splitting algorithm (4.66) and that all the conditions of Proposition 4.22 are satisfied.

(i)&(ii) : Proposition 4.22(iii) asserts that there exists

$$\overline{\overline{x}} = (\overline{x}, \overline{v}_1, \ldots, \overline{v}_m) \in \operatorname{zer}(\widetilde{A} + \widetilde{B}) \tag{4.89}$$

such that $\widetilde{x}_n \rightharpoonup \overline{\overline{x}}$.

(iii)&(iv) : It follows from Proposition 4.22(ii) that $\widetilde{B}\widetilde{x}_n \to \widetilde{B}\overline{\overline{x}}$. Hence, (4.85), (4.87), and (4.89) yield

$$Cx_n \to C\overline{x} \quad \text{and} \quad \big(\forall i \in \{1, \ldots, m\}\big) \quad D_i^{-1}v_{i,n} \to D_i^{-1}\overline{v}_i. \tag{4.90}$$

We derive the results from Definition 4.20 and (i)–(ii) above. ☐

Remark 4.28 Suppose that $(\forall n \in \mathbb{N})$ $\lambda_n \leqslant 1$. Then Corollary 4.27 captures [15, Corollary 6.2].

Acknowledgement. The author thanks his Ph.D. advisor P. L. Combettes for his guidance during this work, which is part of his Ph.D. dissertation.

4.3 Bibliographie

[1] A. ALOTAIBI, P. L. COMBETTES, AND N. SHAHZAD, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set,* SIAM J. Optim., 24 (2014), pp. 2076–2095.

[2] H. ATTOUCH, L. M. BRICEÑO-ARIAS, AND P. L. COMBETTES, *A parallel splitting method for coupled monotone inclusions,* SIAM J. Control Optim., 48 (2010), pp. 3246–3270.

[3] J.-B. BAILLON, R. E. BRUCK, AND S. REICH, *On the asymptotic behavior of nonexpansive mappings and semigroups,* Houston J. Math, 4 (1978), pp. 1–9.

[4] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed., Springer, New York, 2017.

[5] H. BRÉZIS, *Opérateurs Maximaux Monotones et Semi-groupes de Contractions dans les Espaces de Hilbert,* North-Holland/Elsevier, New York, 1973.

[6] L. M. BRICEÑO-ARIAS AND P. L. COMBETTES, *A monotone + skew splitting model for composite monotone inclusions in duality,* SIAM J. Optim., 21 (2011), pp. 1230–1250.

[7] A. CEGIELSKI, *Iterative Methods for Fixed Point Problems in Hilbert Spaces,* Lecture Notes in Mathematics, vol. 2057. Springer, Heidelberg, 2012.

[8] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging,* J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[9] P. L. COMBETTES, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), pp. 475–504.

[10] P. L. COMBETTES, *Monotoned operator theory in convex optimization,* Math. Program., 170 (2018), pp. 177–206.

[11] P. L. COMBETTES AND L. E. GLAUDIN, *Quasi-nonexpansive iterations on the affine hull of orbits : from Mann's mean value algorithm to inertial methods,* SIAM J. Optim., 27 (2017), pp. 2356–2380.

[12] P. L. COMBETTES AND C. L. MÜLLER, *Perspective functions : proximal calculus and applications in high-dimensional statistics,* J. Math. Anal. Appl., 457 (2018), pp. 1283–1306.

[13] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing,* in Fixed-Point Algorithms for Inverse Problems in Science and Engineering, 185–212. Springer, New York (2011)

[14] P. L. COMBETTES AND J.-C. PESQUET, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,* Set-Valued Var. Anal., 20 (2012), pp. 307–330.

[15] P. L. COMBETTES AND B. C. VŨ, *Variable metric quasi-Fejér monotonicity,* Nonlinear Anal., 78 (2013), pp. 17–31.

[16] P. L. COMBETTES AND B. C. VŨ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality,* Optimization, 63 (2014), pp. 1289–1318.

[17] P. L. COMBETTES ANS V. R. WAJS, *Signal recovery by proximal forward-backward splitting,* Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[18] P. L. COMBETTES AND I. YAMADA, *Compositions and convex combinations of averaged nonexpansive operators,* J. Math. Anal. Appl., 425 (2015), pp. 55–70.

[19] J. DUCHI AND Y. SINGER, *Efficient online and batch learning using forward backward splitting,* J. Mach. Learn. Res., 10 (2009), pp. 2899–2934.

[20] R. JENATTON, J. MAIRAL, G. OBOZINSKI, AND F. BACH, *Proximal methods for hierarchical sparse coding,* J. Mach. Learn. Res., 12 (2011), pp. 2297–2334.

[21] J. Peypouquet and S. Sorin, *Evolution equations for maximal monotone operators : asymptotic analysis in continuous and discrete time,* J. Convex Anal., 17 (2010), pp. 1113–1163.

[22] S. Salzo, *The variable metric forward-backward splitting algorithm under mild differentiability assumptions,* SIAM J. Optim., 27 (2017), pp. 2153–2181.

[23] R. E. Showalter, *Monotone Operators in Banach Space and Nonlinear Partial Differential Equations,* Mathematical Surveys and Monographs, vol. 49, Amer. Math. Soc., Providence, RI, 1997.

[24] B. C. Vũ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math., 38 (2013), pp. 667–681.

Chapitre 5

Activation proximale de fonctions lisses dans les méthodes d'éclatement

Les problèmes d'optimisation convexe comprennent typiquement à la fois des fonctions lisses et non lisses. Habituellement, la stratégie est d'utiliser le gradient pour les fonctions lisses et l'opérateur proximal pour les autres. Nous montrons dans ce chapitre qu'il pourrait être judicieux d'essayer de ne pas s'astreindre à cet *a priori*. En effet l'opérateur proximal est calculable pour bon nombre de fonctions lisses. Nous présentons donc de nombreux nouveaux exemples utilisés dans les applications où il est possible de calculer cet opérateur. Un nouveau cadre pour modéliser les problèmes d'admissibilité convexe incompatible est proposé. Finalement, nous appliquons cette stratégie à la restauration d'image, où sont comparées des méthodes complètement implicites et des méthodes implicite/explicite de plusieurs algorithmes d'éclatement. Les résultats obtenus mettent en évidence l'efficacité des activations implicites dans les algorithmes.

5.1 Description et résultats principaux

Le problème étudié dans ce chapitre peut être formalisé comme suit.

Problème 5.1 Soit $\mathcal{H}$ un espace hilbertien réel, soient I et J deux sous-ensembles finis et disjoints de $\mathbb{N}$ tels que $K = I \cup J \neq \varnothing$, soit $f \in \Gamma_0(\mathcal{H})$ et soit $(\mathcal{G}_k)_{k\in K}$ une famille d'espaces hilbertiens réels. Pour tout $k \in K$, soit $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$ et soit $r_k \in \mathcal{G}_k$. Pour tout $i \in I$, soit $g_i \in \Gamma_0(\mathcal{G}_i)$ et pour tout $j \in J$, soit $\mu_j \in \left]0, +\infty\right[$ et soit $h_j \colon \mathcal{G}_j \to \mathbb{R}$ une fonction convexe et différentiable de gradient μ_j-lipschitzien. Supposons que

$$0 \in \operatorname{ran}\left(\partial f + \sum_{i\in I} L_i^* \circ \partial g_i \circ (L_i \cdot - r_i) + \sum_{j\in J} L_j^* \circ (\nabla h_j) \circ (L_j \cdot - r_j)\right). \tag{5.1}$$

L'objectif est de

$$\underset{x\in\mathcal{H}}{\text{minimiser}}\ \ f(x)+\sum_{i\in I} g_i(L_i x-r_i)+\sum_{j\in J} h_j(L_j x-r_j). \tag{5.2}$$

Ce type de problèmes d'éclatement est habituellement étudié d'un point de vue qui offre, pour chaque fonction φ présente dans le modèle, l'alternative suivante :

(i) Si φ est lisse, et on utilise $\nabla\varphi$.

(ii) Sinon on utilise l'opérateur proximal de φ, prox_φ.

Nous allons essayer de nuancer cette méthode de sélection et montrer qu'utiliser l'opérateur proximal, même dans le cas d'une fonction lisse, peut s'avérer efficace. Cette idée d'utiliser des étapes implicites est déjà évoquée dans [11], où il est proposé d'utiliser la version implicite de descente de gradient plutôt que la version explicite, et qu'elle fournirait de meilleurs résultats.

On présente ici deux exemples de calculs d'opérateurs proximaux de fonctions lisses, cruciaux pour notre analyse.

Exemple 5.2 Soit I un sous-ensemble fini non vide de $\mathbb{N}$. Pour tout $i\in I$, soit $\mathcal{G}_i$ un espace hilbertien réel, soit V_i un sous-espace fermé de $\mathcal{G}_i$, soit $r_i\in\mathcal{G}_i$, soit $L_i\in\mathcal{B}(\mathcal{H},\mathcal{G}_i)$ et soit $\alpha_i\in\,]0,+\infty[$. On pose $h\colon\mathcal{H}\to\mathbb{R}\colon x\mapsto(1/2)\sum_{i\in I}\alpha_i d^2_{V_i}(L_i x-r_i)$ et $Q=(\mathrm{Id}+\gamma\sum_{i\in I}\alpha_i L_i^*\text{proj}_{V_i^\perp}L_i)^{-1}$. Soit $\gamma\in\,]0,+\infty[$, on pose $\beta=\sum_{i\in I}\alpha_i\|L_i\|^2$ et soit $x\in\mathcal{H}$. Alors $h\colon\mathcal{H}\to\mathbb{R}$ est convexe et Fréchet différentiable avec un gradient β-lipschitzien,

$$\nabla h(x)=\sum_{i\in I}\alpha_i L_i^*\text{proj}_{V_i^\perp}\big(L_i x-r_i\big)\quad\text{et}\quad \text{prox}_{\gamma h}x=Q\Big(x+\gamma\sum_{i\in I}\alpha_i L_i^*\text{proj}_{V_i^\perp}r_i\Big). \tag{5.3}$$

L'exemple suivant couvre un large spectre d'exemples en régularisant la fonction distance à un ensemble convexe C.

Exemple 5.3 Soit C un sous-ensemble convexe fermé non vide de $\mathcal{H}$, soit $\beta\in\,]0,+\infty[$, soit $\phi\colon\mathbb{R}\to\mathbb{R}$ paire convexe et dérivable avec une dérivée β-lipschitzienne et on pose $h=\phi\circ d_C$. Soient $\gamma\in\,]0,+\infty[$ et $x\in\mathcal{H}$. Alors $h\colon\mathcal{H}\to\mathbb{R}$ est convexe et Fréchet différentiable avec un gradient β-lipschitzien,

$$\nabla h(x)=\begin{cases}\dfrac{\phi'\big(d_C(x)\big)}{d_C(x)}\big(x-\text{proj}_C x\big), & \text{si}\ \ x\notin C;\\ 0, & \text{si}\ \ x\in C\end{cases} \tag{5.4}$$

et

$$\text{prox}_{\gamma h}x=\begin{cases}\text{proj}_C x+\dfrac{\text{prox}_{\gamma\phi}d_C(x)}{d_C(x)}(x-\text{proj}_C x), & \text{si}\ \ x\notin C;\\ x, & \text{si}\ \ x\in C.\end{cases} \tag{5.5}$$

Un autre aspect de ce chapitre est l'étude des problèmes d'admissibilité convexe, couramment étudié en analyse non linéaire. On les formalise comme suit.

Problème 5.4 Soit $\mathcal{H}$ un espace hilbertien réel, soit K un sous-ensemble non vide de $\mathbb{N}$, et soit $(\mathcal{G}_k)_{k\in K}$ une famille d'espaces hilbertiens réels. Pour tout $k \in K$, on suppose que $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$, soit $r_k \in \mathcal{G}_k$ et soit C_k un sous-ensemble convexe fermé non vide de $\mathcal{G}_k$. L'objectif est de

$$\text{trouver } x \in \mathcal{H} \text{ tel que } (\forall k \in K) \ L_k x - r_k \in C_k. \tag{5.6}$$

Malheureusement, il n'est pas rare que cet ensemble de solution soit vide. Il est nécessaire de relaxer le problème et nous proposons un nouveau cadre pour cela. Contrairement à ce qui a pu être fait auparavant, ce nouveau problème permet d'utiliser des contraintes lisses et d'autres non lisses.

Problème 5.5 On considère le cadre du Problème 5.4. Soit (I, J) une partition de K telle que, pour tout $i \in I$, $\phi_i \in \Gamma_0(\mathbb{R})$ une fonction paire qui s'annule seulement en 0 et pour tout $j \in J$, $\psi_j \colon \mathbb{R} \to \mathbb{R}$ une fonction paire différentiable convexe qui s'annule seulement en 0 telle que ψ_j' est lipschitzienne. L'objectif est de

$$\underset{x\in\mathcal{H}}{\text{minimiser}} \ \sum_{i\in I} \phi_i\big(d_{C_i}(L_i x - r_i)\big) + \sum_{j\in J} \psi_j\big(d_{C_j}(L_j x - r_j)\big). \tag{5.7}$$

Il s'agit d'un cas particulier du Problème 5.1 et il couvre de nombreux problèmes importants proposés dans la littérature [14, 21, 25, 37, 56].

5.2 Article en anglais

PROXIMAL ACTIVATION OF SMOOTH FUNCTIONS IN SPLITTING ALGORITHMS FOR CONVEX IMAGE RECOVERY [1]

Abstract. Structured convex optimization problems typically involve a mix of smooth and nonsmooth functions. The common practice is to activate the smooth functions via their gradient and the nonsmooth ones via their proximity operator. We show that, although intuitively natural, this approach is not necessarily the most efficient numerically and that, in particular, activating all the functions proximally may be advantageous. To make this viewpoint viable computationally, we derive a number of new examples of proximity operators of smooth convex functions arising in applications. A

1. P. L. Combettes and L. E. Glaudin, *Proximal activation of smooth functions in splitting algorithms for convex image recovery,* soumis à SIAM J. Imaging Sci.

novel variational model to relax inconsistent convex feasibility problems is also investigated within the proposed framework. Numerical applications to image recovery are presented to compare the behavior of fully proximal versus mixed proximal/gradient implementations of several splitting algorithms.

Keywords. convex optimization, image recovery, inconsistent convex feasibility problem, proximal splitting algorithm, proximity operator

5.2.1 Introduction

Splitting in convex optimization methods for image recovery can be traced back to the influential work of Youla [55, 57]. The convex feasibility framework he proposed consists in formulating the image recovery problem as that of finding an image in a Hilbert space $\mathcal{H}$ satisfying m constraints derived from *a priori* knowledge and the observed data. The constraints are represented by closed convex sets $(C_i)_{1\leqslant i\leqslant m}$ and the problem is therefore to

$$\text{find } x \in S = \bigcap_{i=1}^{m} C_i. \tag{5.8}$$

Now, for every $i \in \{1,\ldots,m\}$, let proj_{C_i} be the projection operator onto C_i, which maps a point $x \in \mathcal{H}$ to the unique closest point in C_i, that is,

$$\text{proj}_{C_i}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y\in\mathcal{H}}{\text{argmin}} \left(\iota_{C_i}(y) + \frac{1}{2}\|x-y\|^2\right), \text{ where } \iota_{C_i}\colon y \mapsto \begin{cases} 0, & \text{if } y \in C_i;\\ +\infty, & \text{if } y \notin C_i. \end{cases} \tag{5.9}$$

The methodology of projection methods is to split the problem of finding a point in S into a sequence of simpler problems involving the sets $(C_i)_{1\leqslant i\leqslant m}$ separately [7, 22]. For instance, the POCS algorithm advocated in [57] is governed by the updating rule

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = (\text{proj}_{C_1} \circ \cdots \circ \text{proj}_{C_m})x_n. \tag{5.10}$$

Projection operators are of limited use beyond feasibility and best approximation problems. Modern image recovery convex variational formulations have a complex structure that requires sophisticated analysis tools and solution methods. To solve such formulations while preserving the spirit of splitting methods in feasibility problems, one approach is to use an extended notion of a projection operator. In [33] it was suggested to use Moreau's notion of a proximity operator [45] for this purpose. Recall that the proximity operator of a proper lower semicontinuous convex function $\varphi\colon \mathcal{H} \to \,]-\infty,+\infty]$ is

$$\text{prox}_{\varphi}\colon \mathcal{H} \to \mathcal{H}\colon x \mapsto \underset{y\in\mathcal{H}}{\text{argmin}} \left(\varphi(y) + \frac{1}{2}\|x-y\|^2\right), \tag{5.11}$$

which reduces to (5.9) when $\varphi = \iota_{C_i}$. We refer the reader to [8, Chapter 24] for a detailed account of the properties of proximity operators with various examples, to [29] for a tutorial on proximal methods in signal processing, and to [5, 10, 18, 20, 28, 46, 47, 48] for specific applications to image recovery. Current proximal splitting methods can handle complex structured convex minimization problems such as the following, which will be the focus of our discussion (see below for notation).

Problem 5.6 Let $\mathcal{H}$ be a real Hilbert space, let I and J be disjoint finite subsets of $\mathbb{N}$ such that $K = I \cup J \neq \varnothing$, let $f \in \Gamma_0(\mathcal{H})$, and let $(\mathcal{G}_k)_{k\in K}$ be a family of real Hilbert spaces. For every $k \in K$, suppose that $0 \neq L_k \in \mathcal{B}(\mathcal{H}, \mathcal{G}_k)$ and let $r_k \in \mathcal{G}_k$. For every $i \in I$, let $g_i \in \Gamma_0(\mathcal{G}_i)$ and, for every $j \in J$, let $\mu_j \in \left]0, +\infty\right[$ and let $h_j \colon \mathcal{G}_j \to \mathbb{R}$ be convex and differentiable with a μ_j-Lipschitzian gradient. Assume that

$$0 \in \operatorname{range}\left(\partial f + \sum_{i\in I} L_i^* \circ \partial g_i \circ (L_i \cdot - r_i) + \sum_{j\in J} L_j^* \circ (\nabla h_j) \circ (L_j \cdot - r_j)\right). \tag{5.12}$$

The goal is to

$$\underset{x\in\mathcal{H}}{\operatorname{minimize}} \;\; f(x) + \sum_{i\in I} g_i(L_i x - r_i) + \sum_{j\in J} h_j(L_j x - r_j). \tag{5.13}$$

The principle of a splitting method for solving (5.13) is to use separately each of the functions f, $(g_i)_{i\in I}$, and $(h_j)_{j\in J}$, and each of the operators $(L_k)_{k\in K}$, so as to reduce the execution of the algorithm to a sequence of simple steps. A prevalent viewpoint in modern proximal splitting algorithms is that to activate each function φ appearing in the model there are two options :

— if φ is smooth, i.e., real-valued and differentiable everywhere with a Lipschitzian gradient, then use $\nabla\varphi$;

— otherwise, use φ proximally, i.e., via its proximity operator (5.11).

In the present paper we propose a more nuanced viewpoint and submit that, when φ is smooth, it may be computationally advantageous to activate it proximally when its proximity operator can be implemented. To motivate this viewpoint, let us first observe that a tight Lipschitz constant for the gradient of φ may not be easy to estimate (see, e.g., [1, 12, 16]), which limits the range of the proximal parameters and may have a detrimental incidence on the speed of convergence. Our second observation is that proximal steps behave numerically quite differently from gradient steps, which may have a positive impact on the asymptotic behavior of the algorithm. On this last score, let us stress that the idea of using proximal steps in smooth problems can already be found in [11, Section 5.8]. There, the problem under consideration was the standard least-squares problem of minimizing the smooth function $\varphi \colon \mathbb{R}^N \to \mathbb{R} \colon x \mapsto \|Ax - b\|^2/2$ in connection with the numerical inversion of the Laplace transform. Given $\gamma \in \left]0, +\infty\right[$ and $x_0 \in \mathbb{R}^N$, the algorithm proposed in [11, Eq. (5.8.3)] assumes the form

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \big(\operatorname{Id} + \gamma A^\top A\big)^{-1}\big(x_n + \gamma A^\top b\big), \tag{5.14}$$

and it is reported to be better than the standard steepest descent approach. This observation echoes the well-known fact that, in numerical analysis, implicit methods have better stability and accuracy properties than explicit ones [52]. Let us note that (5.14) is nothing but an early instance of Martinet's proximal point algorithm [44]

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \operatorname{prox}_{\gamma\varphi} x_n \tag{5.15}$$

for minimizing a function $\varphi \in \Gamma_0(\mathcal{H})$.

The paper is organized as follows. In Section 5.2.2, we enrich the list of known proximity operators by providing new closed form expressions for those of various smooth convex functions commonly encountered in applications. This investigation is of interest in its own right since some splitting algorithms operate exclusively with proximal steps. In connection with the numerical solution of Problem 5.6, we review in Section 5.2.3 some pertinent proximal splitting methods. In Section 5.2.4, we propose a new variational model, based on Problem 5.6 and the results of Section 5.2.2, to relax inconsistent convex feasibility problems. Image recovery applications are presented in Section 5.2.5. Numerical comparisons between splitting algorithms in which smooth functions are activated via gradient steps and those in which all functions are activated via their proximity operators are conducted. While no universal conclusion may be drawn from these experiments, they suggest that fully proximal splitting algorithms deserve to be given serious consideration in applications.

Notation. The notation follows that of [8]. Throughout, $\mathcal{H}$ is a real Hilbert space with scalar product $\langle \cdot \mid \cdot \rangle$, associated norm $\|\cdot\|$, and identity operator Id. Weak convergence is denoted by $\rightharpoonup$. Given a real Hilbert space $\mathcal{G}$, we denote by $\mathcal{B}(\mathcal{H}, \mathcal{G})$ the space of continuous linear operators from $\mathcal{H}$ to $\mathcal{G}$. We set $q = \|\cdot\|^2/2$ and denote by $\Gamma_0(\mathcal{H})$ the class of lower semicontinuous convex functions $f\colon \mathcal{H} \to \left]-\infty, +\infty\right]$ such that $\operatorname{dom} f = \big\{x \in \mathcal{H} \bigm| f(x) < +\infty\big\} \neq \varnothing$. Let $f \in \Gamma_0(\mathcal{H})$. Then f^* denotes the conjugate of f, ∂f the subdifferential of f, and $f \,\square\, g$ the inf-convolution of f and $g \in \Gamma_0(\mathcal{H})$. Let C be a convex subset of $\mathcal{H}$. The strong relative interior of C is denoted by $\operatorname{sri} C$, the indicator function of C by ι_C, the distance function to C by d_C, the support function of C by σ_C and, if C is nonempty and closed, the projection operator onto C by proj_C. The Hilbert direct sum of family of real Hilbert spaces $(\mathcal{H}_i)_{i\in I}$ is denoted by $\bigoplus_{i\in I} \mathcal{H}_i$; in addition if, for every $i \in I$, $f_i\colon \mathcal{H}_i \to [0, +\infty]$, then

$$\bigoplus_{i\in I} f_i\colon \bigoplus_{i\in I} \mathcal{H}_i \to [0, +\infty]\colon (x_i)_{i\in I} \mapsto \sum_{i\in I} f_i(x_i). \tag{5.16}$$

The standard Euclidean norm on $\mathbb{R}^N$ is denoted by $\|\cdot\|_2$.

5.2.2 Proximity operators of smooth convex functions

Let $\beta \in \left]0, +\infty\right[$, let $\gamma \in \left]0, +\infty\right[$, and let $h\colon \mathcal{H} \to \mathbb{R}$ be a convex function with a β-Lipschitzian gradient. Then there exists a function $g \in \Gamma_0(\mathcal{H})$ such that $h = g^* \,\square\, (\beta q)$

[8, Corollary 18.19]. In this case, we derive from [8, Proposition 12.30 and Proposition 24.8(vii)] that

$$\nabla h = \beta\big(\,\mathrm{Id}\, - \mathrm{prox}_{g^*/\beta}\big) \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = \mathrm{Id}\, + \frac{\gamma\beta}{\gamma\beta + 1}\big(\mathrm{prox}_{(\gamma\beta+1)g^*/\beta} - \mathrm{Id}\,\big). \tag{5.17}$$

The closed form expression of $\mathrm{prox}_{\gamma h} x$ above is however of limited use since g^* and its proximity operator are usually not available explicitly. Even when $\mathcal{H} = \mathbb{R}$, computing the proximity operator of a smooth convex function may be involved : for instance the derivative of $h\colon x \mapsto \sqrt[3]{1+x^6}/3$ is $\sqrt[3]{2}$-Lipschitzian but evaluating $\mathrm{prox}_{\gamma h}$ requires solving a high degree polynomial equation. Nonetheless, as we now show, a variety of smooth convex functions encountered in applications have readily computable proximity operators.

Example 5.7 Let I be a nonempty finite set. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let V_i be a closed vector subspace of $\mathcal{G}_i$, let $r_i \in \mathcal{G}_i$, let $L_i \in \mathcal{B}\,(\mathcal{H}, \mathcal{G}_i)$, and let $\alpha_i \in \,]0, +\infty[$. Set $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\sum_{i\in I} \alpha_i d^2_{V_i}(L_i x - r_i)$ and $Q = (\mathrm{Id}\, + \gamma\sum_{i\in I}\alpha_i L_i^* \mathrm{proj}_{V_i^\perp} L_i)^{-1}$. Let $\gamma \in \,]0, +\infty[$, set $\beta = \sum_{i\in I}\alpha_i\|L_i\|^2$, and let $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \sum_{i\in I} \alpha_i L_i^* \mathrm{proj}_{V_i^\perp}\big(L_i x - r_i\big), \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = Q\Bigg(x + \gamma\sum_{i\in I}\alpha_i L_i^* \mathrm{proj}_{V_i^\perp} r_i\Bigg). \tag{5.18}$$

Proof. The convexity of h is clear. We have $h(x) = (1/2)\sum_{i\in I}\alpha_i\|\mathrm{proj}_{V_i^\perp}(L_i x - r_i)\|^2$ and $\nabla h(x)$ is given by (5.18) since $(\forall i \in I)\ \nabla d^2_{V_i}/2 = \mathrm{Id}\, - \mathrm{proj}_{V_i} = \mathrm{proj}_{V_i^\perp}$. Moreover, $(\forall i \in I)$ $\|L_i^*\mathrm{proj}_{V_i^\perp} L_i\| \leqslant \|L_i^*\|\,\|\mathrm{proj}_{V_i^\perp}\|\,\|L_i\| \leqslant \|L_i\|^2$. Hence, for every $y \in \mathcal{H}$,

$$\|\nabla h(x) - \nabla h(y)\| = \Bigg\|\sum_{i\in I}\alpha_i L_i^*\mathrm{proj}_{V_i^\perp} L_i(x-y)\Bigg\| \leqslant \sum_{i\in I}\alpha_i\|L_i\|^2\|x-y\| = \beta\|x-y\|. \tag{5.19}$$

Now set $p = \mathrm{prox}_{\gamma h} x$. Then we derive from (5.18) that

$$x - p = \gamma\nabla h(p) = \gamma\Bigg(\sum_{i\in I}\alpha_i L_i^*\mathrm{proj}_{V_i^\perp} L_i\Bigg)p - \gamma\sum_{i\in I}\alpha_i L_i^*\mathrm{proj}_{V_i^\perp} r_i, \tag{5.20}$$

which yields the expression of $\mathrm{prox}_{\gamma h} x$. □

Example 5.8 Let I be a nonempty finite set. For every $i \in I$, let $\mathcal{G}_i$ be a real Hilbert space, let $r_i \in \mathcal{G}_i$, let $L_i \in \mathcal{B}\,(\mathcal{H}, \mathcal{G}_i)$, and let $\alpha_i \in \,]0, +\infty[$. Set $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto (1/2)\sum_{i\in I}\alpha_i\|L_i x - r_i\|^2$ and $Q = (\mathrm{Id}\, + \gamma\sum_{i\in I}\alpha_i L_i^* L_i)^{-1}$. Let $\gamma \in \,]0, +\infty[$, set $\beta = \sum_{i\in I}\alpha_i\|L_i\|^2$, and let $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \sum_{i\in I}\alpha_i L_i^*\big(L_i x - r_i\big), \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = Q\Bigg(x + \gamma\sum_{i\in I}\alpha_i L_i^* r_i\Bigg). \tag{5.21}$$

Proof. This is a special case of Example 5.7 with $(\forall i \in I)\ V_i = \{0\}$. ☐

The next construction, which involves the distance function d_C to a convex set C, captures a broad range of functions of interest. First, we need the following fact.

Lemma 5.9 [8, Proposition 24.27] *Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be an even convex function that is differentiable on $\mathbb{R} \smallsetminus \{0\}$, set $h = \phi \circ d_C$, and let $x \in \mathcal{H}$. Then*

$$\operatorname{prox}_h x = \begin{cases} x + \dfrac{\operatorname{prox}_{\phi^*} d_C(x)}{d_C(x)}(\operatorname{proj}_C x - x), & \text{if } d_C(x) > \max \partial\phi(0);\\ \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant \max \partial\phi(0). \end{cases} \tag{5.22}$$

Example 5.10 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in \left]0,+\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative, and set $h = \phi \circ d_C$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'\big(d_C(x)\big)}{d_C(x)}\big(x - \operatorname{proj}_C x\big), & \text{if } x \notin C;\\ 0, & \text{if } x \in C, \end{cases} \tag{5.23}$$

and

$$\operatorname{prox}_{\gamma h} x = \begin{cases} \operatorname{proj}_C x + \dfrac{\operatorname{prox}_{\gamma\phi} d_C(x)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } x \notin C;\\ x, & \text{if } x \in C. \end{cases} \tag{5.24}$$

Proof. Since ϕ and d_C are convex and ϕ is increasing on $\left[0,+\infty\right[$, h is convex [8, Proposition 11.7(ii)]. In addition, since 0 is a minimizer of ϕ, $\phi'(0) = 0$ and we derive (5.23) from [8, Proposition 17.33(ii)]. Next, we infer from [8, Proposition 13.26] that $h^* = \sigma_C + \phi^* \circ \|\cdot\|$. Now, we invoke [8, Theorem 18.15] to first deduce that ϕ^* is $(1/\beta)$-strongly convex and that h^* is therefore likewise, and then to conclude that ∇h is β-Lipschitzian. On the other hand, we derive from [8, Proposition 11.7(i)] that 0 is a minimizer of ϕ and therefore that $\partial\phi(0) = \{0\}$. Hence, Lemma 5.9 and [8, Remark 14.4] imply that

$$\begin{aligned} \operatorname{prox}_{\gamma h} x &= \begin{cases} x + \dfrac{\operatorname{prox}_{(\gamma\phi)^*} d_C(x)}{d_C(x)}(\operatorname{proj}_C x - x), & \text{if } d_C(x) > \gamma \max \partial\phi(0);\\ \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant \gamma \max \partial\phi(0) \end{cases}\\ &= \begin{cases} x + \dfrac{d_C(x) - \operatorname{prox}_{\gamma\phi} d_C(x)}{d_C(x)}(\operatorname{proj}_C x - x), & \text{if } d_C(x) > 0;\\ \operatorname{proj}_C x, & \text{if } d_C(x) \leqslant 0 \end{cases}\\ &= \begin{cases} \operatorname{proj}_C x + \dfrac{\operatorname{prox}_{\gamma\phi} d_C(x)}{d_C(x)}(x - \operatorname{proj}_C x), & \text{if } x \notin C;\\ x, & \text{if } x \in C, \end{cases} \end{aligned} \tag{5.25}$$

as announced. □

Example 5.11 Let C be a nonempty closed convex subset of $\mathcal{H}$ and set $h = d_C^2/2$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a nonexpansive gradient, $\nabla f(x) = x - \mathrm{proj}_C x$, and

$$\mathrm{prox}_{\gamma h} x = \frac{1}{\gamma+1}\big(x + \gamma \mathrm{proj}_C x\big). \tag{5.26}$$

Proof. Set $\phi = |\cdot|^2/2$ in Example 5.10. □

Example 5.12 (abstract smooth Vapnik loss function) Let $\varepsilon \in \left]0,+\infty\right[$, let $\beta \in \left]0,+\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative, let $\vartheta\colon \xi \mapsto \max(|\cdot|-\varepsilon, 0)$ be the standard Vapnik loss function, and set $h = \phi\circ\vartheta\circ\|\cdot\|$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'\big(\|x\| - \varepsilon\big)}{\|x\|} x, & \text{if } \|x\| > \varepsilon; \\ 0, & \text{if } \|x\| \leqslant \varepsilon, \end{cases} \tag{5.27}$$

and

$$\mathrm{prox}_{\gamma h} x = \begin{cases} \dfrac{\varepsilon + \mathrm{prox}_{\gamma\phi}(\|x\| - \varepsilon)}{\|x\|} x, & \text{if } \|x\| > \varepsilon; \\ x, & \text{if } \|x\| \leqslant \varepsilon. \end{cases} \tag{5.28}$$

Proof. Let C be the ball with center 0 and radius ε in Example 5.10. □

Here is an extension of the previous example.

Example 5.13 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\varepsilon \in \left]0,+\infty\right[$, let $\beta \in \left]0,+\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative, and set $h = \phi \circ \max(d_C - \varepsilon, 0)$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'\big(d_C(x) - \varepsilon\big)}{d_C(x)} \big(x - \mathrm{proj}_C x\big), & \text{if } d_C(x) > \varepsilon; \\ 0, & \text{if } d_C(x) \leqslant \varepsilon, \end{cases} \tag{5.29}$$

and

$$\mathrm{prox}_{\gamma h} x = \begin{cases} \mathrm{proj}_C x + \dfrac{\varepsilon + \mathrm{prox}_{\gamma\phi}(d_C(x) - \varepsilon)}{d_C(x)} (x - \mathrm{proj}_C x), & \text{if } d_C(x) > \varepsilon; \\ x, & \text{if } d_C(x) \leqslant \varepsilon. \end{cases} \tag{5.30}$$

Proof. Set $\psi = \phi \circ \max(|\cdot| - \varepsilon, 0)$, apply Example 5.10 in $\mathbb{R}$ to ψ, and use Example 5.12 for the gradient and the proximity operator of ψ. □

Example 5.14 (abstract Huber function) Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\rho \in\,]0, +\infty[$, and set

$$h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto \begin{cases} \rho d_C(x) - \dfrac{\rho^2}{2}, & \text{if } d_C(x) > \rho; \\ \dfrac{d_C(x)^2}{2}, & \text{if } d_C(x) \leqslant \rho. \end{cases} \tag{5.31}$$

Let $\gamma \in\,]0, +\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a nonexpansive gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\rho}{d_C(x)}(x - \text{proj}_C x), & \text{if } d_C(x) > \rho; \\ x - \text{proj}_C x, & \text{if } d_C(x) \leqslant \rho, \end{cases} \tag{5.32}$$

and

$$\text{prox}_{\gamma h} x = \begin{cases} x + \dfrac{\gamma\rho}{d_C(x)}(\text{proj}_C x - x), & \text{if } d_C(x) > (\gamma+1)\rho; \\ \dfrac{1}{\gamma+1}(x + \gamma \text{proj}_C x), & \text{if } d_C(x) \leqslant (\gamma+1)\rho. \end{cases} \tag{5.33}$$

Proof. Let

$$\phi\colon \mathbb{R} \to \mathbb{R}\colon \xi \mapsto \begin{cases} \rho|\xi| - \dfrac{\rho^2}{2}, & \text{if } |\xi| > \rho; \\ \dfrac{|\xi|^2}{2}, & \text{if } |\xi| \leqslant \rho \end{cases} \tag{5.34}$$

be the standard Huber function with parameter ρ. Then ϕ' is 1-Lipschitzian. In addition, using the expression of $\text{prox}_{\gamma\phi}$ from [8, Example 24.9] and then Example 5.10, we obtain (5.33). □

Example 5.15 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\omega \in\,]0, +\infty[$, and set $\beta = \omega^2$ and $h = \omega d_C - \ln(1 + \omega d_C)$. Let $\gamma \in\,]0, +\infty[$ and let $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient, and

$$\text{prox}_{\gamma h} x = \begin{cases} \text{proj}_C x + \dfrac{\gamma\omega^2 + 1 - \omega d_C(x) - \sqrt{|\omega d_C(x) - \gamma\omega^2 - 1|^2 + 4\omega d_C(x)}}{2\omega d_C(x)}(\text{proj}_C x - x), & \text{if } x \notin C; \\ x, & \text{if } x \in C. \end{cases} \tag{5.35}$$

Proof. Set $\phi = \omega|\cdot| - \ln(1+\omega|\cdot|)$. Then ϕ' is ω^2-Lipschitzian. Furthermore, we derive $\text{prox}_{\gamma\phi}$ by arguing as in [8, Example 24.42] (where $\gamma = 1$) and then invoking Example 5.10. □

Example 5.16 Let $\beta \in \left]0,+\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative, and set $h = \phi \circ \|\cdot\|$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \begin{cases} \dfrac{\phi'(\|x\|)}{\|x\|}x, & \text{if } x \neq 0;\\ 0, & \text{if } x = 0, \end{cases} \quad\text{and}\quad \text{prox}_{\gamma h}x = \begin{cases} \dfrac{\text{prox}_{\gamma\phi}\|x\|}{\|x\|}x, & \text{if } x \neq 0;\\ x, & \text{if } x = 0. \end{cases} \tag{5.36}$$

Proof. Set $C = \{0\}$ in Example 5.10. □

The following extension of Example 5.10 involves a composition with a linear operator.

Example 5.17 Let $\mathcal{G}$ be a real Hilbert space and let $M \in \mathcal{B}(\mathcal{H},\mathcal{G})$ be such that $MM^* = \theta\,\mathrm{Id}$ for some $\theta \in \left]0,+\infty\right[$. Let D be a nonempty closed convex subset of $\mathcal{G}$, let $\mu \in \left]0,+\infty\right[$, let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a μ-Lipschitzian derivative, and set $h = \phi \circ d_D \circ M$ and $\beta = \mu\|M\|^2$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient, and

$$\text{prox}_{\gamma h}x = \begin{cases} x + \dfrac{\theta^{-1}\big(d_D(Mx) - \text{prox}_{\gamma\theta\phi}d_D(Mx)\big)}{d_D(Mx)}M^*\big(\text{proj}_D(Mx) - Mx\big), & \text{if } Mx \notin D;\\ x, & \text{if } Mx \in D. \end{cases} \tag{5.37}$$

Proof. Set $g = \gamma\phi \circ d_D$. We first derive from [8, Proposition 24.14] that

$$\text{prox}_{\gamma h}x = \text{prox}_{g\circ M}x = x + \theta^{-1}M^*\big(\text{prox}_{\theta g}(Mx) - Mx\big). \tag{5.38}$$

We then obtain the expression of $\text{prox}_{\theta g} = \text{prox}_{\gamma\theta\phi\circ d_D}$ from Example 5.10 and, in turn, (5.37). □

Remark 5.18 The condition $MM^* = \theta\,\mathrm{Id}$ used in Example 5.17 arises in particular in problems involving tight frame representations [19]. When it is not satisfied, one can still deal with smooth functions of the type $\phi\circ d_C\circ M$ in modern structured proximal splitting techniques by activating $\text{prox}_{\phi\circ d_C}$ and M separately; see Propositions 5.34, 5.36, and 5.38 below and [8]. One can then invoke Example 5.10 to compute the former.

Example 5.19 Let $(\Omega,\mathcal{F},\mu)$ be a complete σ-finite measure space, let $(\mathsf{H},\langle\cdot\mid\cdot\rangle_{\mathsf{H}})$ be a separable real Hilbert space, let C be a closed convex subset of H such that $0 \in \mathsf{C}$, and let $\phi\colon \mathbb{R} \to \mathbb{R}$ be even, convex, and differentiable with a β-Lipschitzian derivative. Suppose that $\mathcal{H} = L^2((\Omega,\mathcal{F},\mu);\mathsf{H})$, and that $\mu(\Omega) < +\infty$ or $\phi(0) = 0$. Set $h\colon \mathcal{H} \to \mathbb{R}\colon x \mapsto$

$\int_\Omega \phi(d_C(x(\omega)))\mu(d\omega)$. Let $\gamma \in \left]0,+\infty\right[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient, and for μ-almost every $w \in \Omega$,

$$\big(\nabla h(x)\big)(\omega) = \begin{cases} \dfrac{\phi'(d_C(x(\omega)))}{d_C(x(\omega))}\big(x(\omega) - \mathrm{proj}_C x(\omega)\big), & \text{if } \; x(\omega) \notin C; \\ 0, & \text{if } \; x(\omega) \in C \end{cases} \tag{5.39}$$

and

$$\big(\mathrm{prox}_{\gamma h} x\big)(\omega) = \begin{cases} \mathrm{proj}_C x(\omega) + \dfrac{\mathrm{prox}_{\gamma\phi} d_C(x(\omega))}{d_C(x(\omega))}\big(x(\omega) - \mathrm{proj}_C x(\omega)\big), & \text{if } \; x(\omega) \notin C; \\ x(\omega), & \text{if } \; x(\omega) \in C. \end{cases} \tag{5.40}$$

Proof. Set $\varphi = \phi \circ d_C$. As seen in Example 5.10, $\varphi\colon \mathsf{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \nabla\varphi(\mathsf{x}) = \begin{cases} \dfrac{\phi'\big(d_C(\mathsf{x})\big)}{d_C(\mathsf{x})}\big(\mathsf{x} - \mathrm{proj}_C \mathsf{x}\big), & \text{if } \; \mathsf{x} \notin C; \\ 0, & \text{if } \; \mathsf{x} \in C, \end{cases} \tag{5.41}$$

and

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \mathrm{prox}_{\gamma\varphi} \mathsf{x} = \begin{cases} \mathrm{proj}_C \mathsf{x} + \dfrac{\mathrm{prox}_{\gamma\phi} d_C(\mathsf{x})}{d_C(\mathsf{x})}(\mathsf{x} - \mathrm{proj}_C \mathsf{x}), & \text{if } \; \mathsf{x} \notin C; \\ \mathsf{x}, & \text{if } \; \mathsf{x} \in C. \end{cases} \tag{5.42}$$

Since ϕ is convex and even, it is minimized by 0 [8, Proposition 11.7(i)], and we have $\phi'(0) = 0$ and, since $0 \in C$, $\varphi(0) = \phi(0)$, while (5.41) yields $\nabla\varphi(0) = 0$. Consequently, by virtue of the descent lemma [8, Theorem 18.15(iii)],

$$(\forall \mathsf{x} \in \mathsf{H}) \quad \varphi(\mathsf{x}) \leqslant \varphi(0) + \langle \mathsf{x} \mid \nabla\varphi(0)\rangle_{\mathsf{H}} + \frac{\beta}{2}\|\mathsf{x}\|_{\mathsf{H}}^2 = \varphi(0) + \frac{\beta}{2}\|\mathsf{x}\|_{\mathsf{H}}^2. \tag{5.43}$$

Hence,

$$h(x) = \int_\Omega \varphi(x(\omega))\mu(d\omega) \leqslant \varphi(0)\mu(\Omega) + \frac{\beta}{2}\|x\|^2 < +\infty. \tag{5.44}$$

On the other hand, [8, Proposition 16.63(ii)] asserts that $\nabla h(x) = (\nabla\varphi) \circ x$ μ-a.e. which, combined with (5.41), yields (5.39). Now let y be in $\mathcal{H}$. Then

$$\begin{aligned} \|\nabla h(x) - \nabla h(y)\|^2 &= \int_\Omega \|(\nabla h(x))(\omega) - (\nabla h(y))(\omega)\|_{\mathsf{H}}^2 \mu(d\omega) \\ &= \int_\Omega \|\nabla\varphi(x(\omega)) - \nabla\varphi(y(\omega))\|_{\mathsf{H}}^2 \mu(d\omega) \\ &\leqslant \beta^2 \int_\Omega \|x(\omega) - y(\omega)\|_{\mathsf{H}}^2 \mu(d\omega) \\ &= \beta^2\|x - y\|^2, \end{aligned} \tag{5.45}$$

which shows that ∇h is β-Lipschitzian. Finally, we apply [8, Proposition 24.13] to derive (5.40) from (5.42). □

Remark 5.20 Let Ω be a nonempty bounded smooth open subset of $\mathbb{R}^2$, let $\mathsf{H} = \mathbb{R}^2$, let μ be the Lebesgue measure, and suppose that $\mathsf{C} = \{0\}$ in Example 5.19. Furthermore, let ϕ be the Huber function of (5.34) and, for every $x \in H_0^1(\Omega)$, let Dx be the gradient of x. Then the function

$$h \circ D\colon H_0^1(\Omega) \to \mathbb{R}\colon x \mapsto \int_\Omega \phi(\|Dx(\omega)\|_2) d\omega \tag{5.46}$$

can be found in [39, 41, 43, 49] and it is called the Gauss-TV (or TV-Huber) function.

We now turn to a different scheme to construct smooth convex functions.

Example 5.21 Let $\varphi \in \Gamma_0(\mathcal{H})$, let $\beta \in \,]0,+\infty[$, and define $h = \beta q - (\beta q)\,\square\,\varphi$. Let $\gamma \in \,]0,+\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \beta \mathrm{prox}_{\varphi/\beta} x, \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = x - \beta\gamma \mathrm{prox}_{\frac{\varphi}{\beta(1+\beta\gamma)}}\left(\frac{1}{1+\beta\gamma}x\right). \tag{5.47}$$

Proof. It follows from [8, Proposition 12.30] that $\nabla h = \beta(\mathrm{Id} - (\mathrm{Id} - \mathrm{prox}_{\varphi/\beta})) = \beta \mathrm{prox}_{\varphi/\beta}$, which is β-Lipschitzian since $\mathrm{prox}_{\varphi/\beta}$ is nonexpansive [8, Proposition 12.28]. Finally (5.47) follows from [8, Proposition 24.8(viii)]. □

Example 5.22 Let C be a nonempty closed convex subset of $\mathcal{H}$, let $\beta \in \,]0,+\infty[$, and set $h = \beta(q - d_C^2/2)$. Let $\gamma \in \,]0,+\infty[$ and $x \in \mathcal{H}$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x) = \beta \mathrm{proj}_C x, \quad \text{and} \quad \mathrm{prox}_{\gamma h} x = x - \beta\gamma \mathrm{proj}_C\left(\frac{1}{1+\beta\gamma}x\right). \tag{5.48}$$

Proof. Set $\varphi = \iota_C$ in Example 5.21. □

Remark 5.23 For $\beta = 1$, the function h in Example 5.22 is the generalized Huber function of [24, Example 3.2], that is

$$(\forall x \in \mathcal{H}) \quad h(x) = \begin{cases} \langle x \mid \mathrm{proj}_C x\rangle - \dfrac{\|\mathrm{proj}_C x\|^2}{2}, & \text{if } x \notin C; \\ \dfrac{\|x\|^2}{2}, & \text{if } x \in C. \end{cases} \tag{5.49}$$

If $\mathcal{H} = \mathbb{R}$ and $C = [-\rho, \rho]$ for some $\rho \in \,]0,+\infty[$, h reduces to the standard Huber function (5.34).

The next example revisits a construction proposed in [33] ; see also [31, 35, 36] for special cases.

Example 5.24 Suppose that $\mathcal{H}$ is separable and that $\varnothing \neq \mathbb{K} \subset \mathbb{N}$, and let $(e_k)_{k\in\mathbb{K}}$ be an orthonormal basis of $\mathcal{H}$. For every $k \in \mathbb{K}$, let $\beta_k \in \left]0,+\infty\right[$ and let $\phi_k \colon \mathbb{R} \to \mathbb{R}$ be a differentiable convex function such that $\phi_k \geqslant \phi_k(0) = 0$ and ϕ'_k is β_k-Lipschitzian. Suppose that $\beta = \sup_{k\in\mathbb{K}} \beta_k < +\infty$ and define $(\forall x \in \mathcal{H})\ h(x) = \sum_{k\in\mathbb{K}} \phi_k(\langle x \mid e_k\rangle)$. Let $\gamma \in \left]0,+\infty\right[$. Then $h\colon \mathcal{H} \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$(\forall x \in \mathcal{H}) \quad \nabla h(x) = \sum_{k\in\mathbb{K}} \phi'_k(\langle x \mid e_k\rangle)e_k, \tag{5.50}$$

and

$$(\forall x \in \mathcal{H}) \quad \operatorname{prox}_{\gamma h} x = \sum_{k\in\mathbb{K}} \big(\operatorname{prox}_{\gamma\phi_k} \langle x \mid e_k\rangle\big)e_k. \tag{5.51}$$

Proof. The identity (5.51) is established in [8, Proposition 24.16]. The frame analysis operator is

$$F\colon \mathcal{H} \to \ell^2(\mathbb{K})\colon x \mapsto (\langle x \mid e_k\rangle)_{k\in\mathbb{K}} \tag{5.52}$$

and its adjoint is the frame synthesis operator

$$F^*\colon \ell^2(\mathbb{K}) \to \mathcal{H}\colon (\xi_k)_{k\in\mathbb{K}} \mapsto \sum_{k\in\mathbb{K}} \xi_k e_k. \tag{5.53}$$

Now denote by $\mathsf{x} = (\xi_k)_{k\in\mathbb{K}}$ a generic element in $\ell^2(\mathbb{K})$ and define

$$\varphi\colon \ell^2(\mathbb{K}) \to \left]-\infty,+\infty\right] : \mathsf{x} \mapsto \sum_{k\in\mathbb{K}} \phi_k(\xi_k). \tag{5.54}$$

Then $h = \varphi \circ F$. Since all the functions $(\phi_k)_{k\in\mathbb{K}}$ are minimized at 0, we have $(\forall k \in \mathbb{K})$ $\phi'_k(0) = 0 = \phi_k(0)$. In turn, we derive from the descent lemma [8, Theorem 18.15(iii)] that

$$(\forall k \in \mathbb{K})(\forall \xi_k \in \mathbb{R}) \quad \phi_k(\xi_k) \leqslant \phi_k(0) + (\xi_k - 0)\phi'_k(0) + \frac{\beta_k}{2}|0 - \xi_k|^2 = \frac{\beta}{2}|\xi_k|^2. \tag{5.55}$$

As a result,

$$\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big) \quad \varphi(\mathsf{x}) = \sum_{k\in\mathbb{K}} \phi_k(\xi_k) \leqslant \frac{\beta}{2}\sum_{k\in\mathbb{K}} |\xi_k|^2 = \frac{\beta}{2}\|\mathsf{x}\|^2 \tag{5.56}$$

and, therefore, $\varphi\colon \ell^2(\mathbb{K}) \to \mathbb{R}$. In addition,

$$(\forall k \in \mathbb{K})(\forall \xi_k \in \mathbb{R}) \quad |\phi'_k(\xi_k)|^2 = |\phi'_k(\xi_k) - \phi'_k(0)|^2 \leqslant \beta_k^2|\xi_k - 0|^2 \leqslant \beta^2|\xi_k|^2, \tag{5.57}$$

which yields

$$\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big) \quad \sum_{k\in\mathbb{K}} |\phi_k'(\xi_k)|^2 \leqslant \beta^2 \sum_{k\in\mathbb{K}} |\xi_k|^2 = \beta^2 \|\mathsf{x}\|^2 < +\infty, \tag{5.58}$$

and allows us to conclude that φ is differentiable with $(\forall \mathsf{x} \in \ell^2(\mathbb{K}))\ \nabla\varphi(\mathsf{x}) = (\phi_k'(\xi_k))_{k\in\mathbb{K}}$. Moreover,

$$\begin{aligned}\big(\forall \mathsf{x} \in \ell^2(\mathbb{K})\big)\big(\forall \mathsf{y} \in \ell^2(\mathbb{K})\big) \quad \|\nabla\varphi(\mathsf{x}) - \nabla\varphi(\mathsf{y})\|^2 &= \sum_{k\in\mathbb{K}} |\phi_k'(\xi_k) - \phi_k'(\eta_k)|^2 \\ &\leqslant \sum_{k\in\mathbb{K}} \beta_k^2 |\xi_k - \eta_k|^2 \\ &\leqslant \beta^2 \|\mathsf{x} - \mathsf{y}\|^2. \end{aligned} \tag{5.59}$$

Altogether, $\varphi\colon \ell^2(\mathbb{K}) \to \mathbb{R}$ is convex and differentiable with a β-Lipschitzian gradient. Hence $h = \varphi \circ F\colon \mathcal{H} \to \mathbb{R}$ is convex and differentiable, with $\nabla h = F^* \circ \nabla\varphi \circ F$. Furthermore, since F and F^* are isometries,

$$\begin{aligned}(\forall x \in \mathcal{H})(\forall y \in \mathcal{H}) \quad \|\nabla h(x) - \nabla h(y)\| &= \|F^*(\nabla\varphi(Fx)) - F^*(\nabla\varphi(Fy))\| \\ &= \|\nabla\varphi(Fx) - \nabla\varphi(Fy)\| \\ &\leqslant \beta\|Fx - Fy\| \\ &= \beta\|x - y\|, \end{aligned} \tag{5.60}$$

which shows that ∇h is β-Lipschitzian. □

Remark 5.25 It is clear from the proof of Example 5.24 that if the condition

$$\phi_k \geqslant \phi_k(0) = 0 \tag{5.61}$$

is not satisfied for a finite number of indices in $k \in \mathbb{K}$, the results remain valid. In particular, if $\mathcal{H}$ is finite-dimensional, (5.61) is not required. An example of a smooth convex function $\phi_k\colon \mathbb{R} \to \mathbb{R}$ with explicit proximity operator is $\phi_k = d_{C_k}^2/2$, where C_k is a nonempty closed interval in $\mathbb{R}$ (see Example 5.11). In this case, (5.61) holds if and only if $0 \in C_k$. If $C_k = [1, +\infty[$, ϕ_k is the squared hinge loss [50]; if $C_k = [-\varepsilon, \varepsilon]$ for some $\varepsilon \in\,]0, +\infty[$, ϕ_k is the smooth Vapnik insensitive loss [4] (see also Example 5.12). 2cSpecializing Examples 5.10, 5.14, 5.15, and 5.22 to $\mathcal{H} = \mathbb{R}$ provides further examples of functions of interest. For instance, taking $\phi_k = \psi_k \circ d_{C_k}$, where ψ_k is the Huber function (5.34) and $C_k = [1, +\infty[$ yields the modified Huber function [58].

Example 5.26 Let M be a strictly positive integer and let $\varnothing \neq \mathbb{K} \subset \mathbb{N}$. For every $i \in \{1, \ldots, M\}$, suppose that $\mathcal{H}_i$ is a separable real Hilbert space with orthonormal basis $(e_{i,k})_{k\in\mathbb{K}}$. Let $\beta \in\,]0, +\infty[$, and let $\phi\colon \mathbb{R} \to \mathbb{R}$ be a differentiable convex function such that $\phi \geqslant \phi(0) = 0$ and ϕ' is β-Lipschitzian. For every $(x_1, \ldots, x_M)$

in $\mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M$, set $h(x_1,\ldots,x_M) = \sum_{k\in\mathbb{K}} \phi(\|(\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\|_2)$. Let $\gamma \in \left]0,+\infty\right[$, let $(x_1,\ldots,x_M) \in \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M$, and set, for every $k \in \mathbb{K}$,

$$\alpha_k = \begin{cases} \dfrac{\phi'(\|(\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\|_2)}{\|(\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\|_2}, & \text{if } \max\limits_{1\leqslant i\leqslant M} |\langle x_i \mid e_{i,k}\rangle| > 0;\\ 0, & \text{otherwise,} \end{cases} \tag{5.62}$$

and

$$\delta_k = \begin{cases} \dfrac{\operatorname{prox}_{\gamma\phi}(\|(\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\|_2)}{\|(\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\|_2}, & \text{if } \max\limits_{1\leqslant i\leqslant M} |\langle x_i \mid e_{i,k}\rangle| > 0;\\ 1, & \text{otherwise.} \end{cases} \tag{5.63}$$

Then $h\colon \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M \to \mathbb{R}$ is convex and Fréchet differentiable with a β-Lipschitzian gradient,

$$\nabla h(x_1,\ldots,x_M) = \left(\sum_{k\in\mathbb{K}} \alpha_k \langle x_1 \mid e_{1,k}\rangle e_{1,k},\ldots,\sum_{k\in\mathbb{K}} \alpha_k \langle x_M \mid e_{M,k}\rangle e_{M,k}\right), \tag{5.64}$$

and

$$\operatorname{prox}_{\gamma h}(x_1,\ldots,x_M) = \left(\sum_{k\in\mathbb{K}} \delta_k \langle x_1 \mid e_{1,k}\rangle e_{1,k},\ldots,\sum_{k\in\mathbb{K}} \delta_k \langle x_M \mid e_{M,k}\rangle e_{M,k}\right). \tag{5.65}$$

Proof. Arguing as in (5.55)–(5.56), we obtain that h is real-valued and, arguing as in (5.57), we get $\sup_{k\in\mathbb{K}} \alpha_k \leqslant \beta$. Hence $(\forall i \in \{1,\ldots,M\})$ $(\alpha_k\langle x_i \mid e_{i,k}\rangle)_{k\in\mathbb{K}} \in \ell^2(\mathbb{K})$ and $\sum_{k\in\mathbb{K}} \alpha_k \langle x_i \mid e_{i,k}\rangle e_{i,k} \in \mathcal{H}_i$. Likewise, since ϕ is minimized at 0, $\operatorname{prox}_{\gamma\phi} 0 = 0$ and thus $\sup_{k\in\mathbb{K}} \delta_k \leqslant 1$. It therefore follows that $(\forall i \in \{1,\ldots,M\})$ $(\delta_k\langle x_i \mid e_{i,k}\rangle)_{k\in\mathbb{K}} \in \ell^2(\mathbb{K})$ and $\sum_{k\in\mathbb{K}} \delta_k \langle x_i \mid e_{i,k}\rangle e_{i,k} \in \mathcal{H}_i$. Now let $\ell^2(\mathbb{K};\mathbb{R}^M) = \bigoplus_{k\in\mathbb{K}} \mathbb{R}^M$ be the space of square summable sequences with entries in $\mathbb{R}^M$, and set

$$\begin{cases} \varphi\colon \mathbb{R}^M \to \mathbb{R}\colon (\xi_1,\ldots,\xi_M) \mapsto \phi(\|(\xi_1,\ldots,\xi_M)\|_2)\\ U\colon \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M \to \ell^2(\mathbb{K};\mathbb{R}^M)\colon (x_1,\ldots,x_M) \mapsto \big((\langle x_1 \mid e_{1,k}\rangle,\ldots,\langle x_M \mid e_{M,k}\rangle)\big)_{k\in\mathbb{K}}. \end{cases} \tag{5.66}$$

Then U is a bijective isometry,

$$U^{-1} = U^*\colon \ell^2(\mathbb{K};\mathbb{R}^M) \to \mathcal{H}_1 \oplus \cdots \oplus \mathcal{H}_M\colon \tag{5.67}$$

$$\big((\mu_{1,k},\ldots,\mu_{M,k})\big)_{k\in\mathbb{K}} \mapsto \left(\sum_{k\in\mathbb{K}} \mu_{1,k} e_{1,k},\ldots,\sum_{k\in\mathbb{K}} \mu_{M,k} e_{M,k}\right), \tag{5.68}$$

and $h = \big(\bigoplus_{k\in\mathbb{K}} \varphi\big) \circ U$. In turn,

$$\nabla h = U^*\circ\nabla\left(\bigoplus_{k\in\mathbb{K}} \varphi\right)\circ U = U^*\circ\nabla\left(\bigoplus_{k\in\mathbb{K}} (\phi\circ\|\cdot\|_2)\right)\circ U = U^*\circ\left(\bigotimes_{k\in\mathbb{K}} \nabla(\phi\circ\|\cdot\|_2)\right)\circ U, \tag{5.69}$$

which yields (5.64) using Example 5.16. On the other hand, using [8, Proposition 24.14], we obtain

$$\operatorname{prox}_{\gamma h} = \operatorname{prox}_{\gamma(\bigoplus_{k\in\mathbb{K}} \varphi)\circ U} = \left(U^* \circ \operatorname{prox}_{\gamma \bigoplus_{k\in\mathbb{K}} \varphi} \circ U\right) = U^* \circ \left(\bigotimes_{k\in\mathbb{K}} \operatorname{prox}_{\gamma\phi\circ\|\cdot\|_2}\right) \circ U, \tag{5.70}$$

which yields (5.65) using Example 5.16. □

5.2.3 Splitting algorithms

In this section, we review several proximal splitting methods which are relevant to our discussion and will be used in our numerical comparisons (see [8] for the supporting theory). We start with the forward-backward splitting algorithm.

Algorithm 5.27 (forward-backward) Let $\beta \in \left]0, +\infty\right[$, let $f \in \Gamma_0(\mathcal{H})$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a β-Lipschitzian gradient, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $\left]0, 2/\beta\right[$ such that $0 < \inf_{n\in\mathbb{N}} \gamma_n \leqslant \sup_{n\in\mathbb{N}} \gamma_n < 2/\beta$. Let $x_0 \in \mathcal{H}$ and iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} y_n = x_n - \gamma_n \nabla g(x_n) \\ x_{n+1} = \operatorname{prox}_{\gamma_n f} y_n. \end{array} \right. \end{array} \tag{5.71}$$

Proposition 5.28 [31] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.27 and suppose that* $\operatorname{Argmin}(f+g) \neq \varnothing$. *Then there exists $x \in$* $\operatorname{Argmin}(f+g)$ *such that $x_n \rightharpoonup x$. In addition* $(f+g)(x_n) - \inf(f+g)(\mathcal{H}) = o(1/n)$.

Inertial variants of the above method have been popularized by [10]. They require additional storage capabilities but have been shown to be advantageous in terms of convergence speed in certain situations. The following implementation proposed in [17] guarantees convergence of the iterates.

Algorithm 5.29 (inertial forward-backward) Let $\beta \in \left]0, +\infty\right[$, let $f \in \Gamma_0(\mathcal{H})$, let $g\colon \mathcal{H} \to \mathbb{R}$ be convex and differentiable with a β-Lipschitzian gradient, let $\gamma \in \left]0, 1/\beta\right]$, and let $\alpha \in \left]2, +\infty\right[$. Let $x_0 = x_{-1} \in \mathcal{H}$ and iterate

$$\begin{array}{l} \text{for } n = 0, 1, \ldots \\ \left\lfloor \begin{array}{l} z_n = x_n + \dfrac{n-1}{n+\alpha}(x_n - x_{n-1}) \\ y_n = z_n - \gamma \nabla g(z_n) \\ x_{n+1} = \operatorname{prox}_{\gamma f} y_n. \end{array} \right. \end{array} \tag{5.72}$$

Proposition 5.30 [17] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.29 and suppose that* $\operatorname{Argmin}(f+g) \neq \varnothing$. *Then there exists $x \in$* $\operatorname{Argmin}(f+g)$ *such that $x_n \rightharpoonup x$. In addition* $(f+g)(x_n) - \inf(f+g)(\mathcal{H}) = O(1/n^2)$.

The next algorithm does not require smoothness of any of the functions.

Algorithm 5.31 (Douglas-Rachford) Let f and g be functions in $\Gamma_0(\mathcal{H})$ such that $0 \in \operatorname{sri}(\operatorname{dom} g - \operatorname{dom} f)$, let $\gamma \in \left]0,+\infty\right[$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be sequence in $[0,2]$ such that $\sum_{n\in\mathbb{N}} \lambda_n(2-\lambda_n) = +\infty$. Let $y_0 \in \mathcal{H}$ and iterate

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
z_n = \operatorname{prox}_{\gamma g} y_n \\
x_n = \operatorname{prox}_{\gamma f}(2z_n - y_n) \\
y_{n+1} = y_n + \lambda_n(x_n - z_n).
\end{array}
\right.
\end{array}
\tag{5.73}
$$

Proposition 5.32 [28] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.31 and suppose that* $\operatorname{Argmin}(f+g) \neq \varnothing$*. Then there exists* $x \in \operatorname{Argmin}(f+g)$ *such that* $x_n \rightharpoonup x$.

Although this feature will not be used in Section 5.2.5, it should be noted that the forward-backward [31], inertial forward-backward [6], and Douglas-Rachford [23] algorithms tolerate errors in the implementations of the proximity operators. The next three algorithms are specifically tailored to handle Problem 5.6 and they also compute dual solutions. For brevity, we present only the primal convergence result and the error-free, unrelaxed formulations of these algorithms. The first one is known as the primal-dual forward-backward-forward algorithm; see [30] for details and [54] for a variable metric version.

Algorithm 5.33 Consider the setting of Problem 5.6. Set $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2} + \sum_{j\in J}\mu_j\|L_j\|^2$, let $\varepsilon \in \left]0,1/(\beta+1)\right[$, let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $[\varepsilon,(1-\varepsilon)/\beta]$, and let $(\forall i \in I)\ v_{i,0}^* \in \mathcal{G}_i$. Let $x_0 \in \mathcal{H}$ and iterate

$$
\begin{array}{l}
\text{for } n = 0,1,\ldots \\
\left\lfloor
\begin{array}{l}
y_{1,n} = x_n - \gamma_n\Big(\sum_{i\in I} L_i^* v_{i,n}^* + \sum_{j\in J} L_j^*\big(\nabla h_j(L_j x_n - r_j)\big)\Big) \\
p_{1,n} = \operatorname{prox}_{\gamma_n f} y_{1,n} \\
\text{for every } i \in I \\
\left\lfloor
\begin{array}{l}
y_{2,i,n} = v_{i,n}^* + \gamma_n L_i x_n \\
p_{2,i,n} = \operatorname{prox}_{\gamma_n g_i^*}(y_{2,i,n} - \gamma_n r_i) \\
q_{2,i,n} = p_{2,i,n} + \gamma_n L_i p_{1,n} \\
v_{i,n+1}^* = v_{i,n}^* - y_{2,i,n} + q_{2,i,n}
\end{array}
\right. \\
q_{1,n} = p_{1,n} - \gamma_n\Big(\sum_{i\in I} L_i^* p_{2,i,n} + \sum_{j\in J} L_j^*\big(\nabla h_j(L_j p_{1,n} - r_j)\big)\Big) \\
x_{n+1} = x_n - y_{1,n} + q_{1,n}.
\end{array}
\right.
\end{array}
\tag{5.74}
$$

Proposition 5.34 [30] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.33. Then there exists a solution x to (5.13) such that $x_n \rightharpoonup x$.*

The next algorithm is an implementation of the forward-backward algorithm in a renormed primal-dual space (see [18, 26, 34, 38, 53] for special cases and variants, and [32] for a more general variable metric version).

Algorithm 5.35 Consider the setting of Problem 5.6 and let $(\tau_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $(\forall n\in\mathbb{N})\ \tau_{n+1}\geqslant\tau_n$. For every $i\in I$, let $v_{i,0}^*\in\mathcal{G}_i$, let $(\sigma_{i,n})_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $(\forall n\in\mathbb{N})\ \sigma_{i,n+1}\geqslant\sigma_{i,n}$. Suppose that

$$\sup_{n\in\mathbb{N}}\left(\sqrt{\tau_n\sum_{i\in I}\sigma_{i,n}\|L_i\|^2}+\frac{1}{2}\max\left\{\tau_n,\max_{i\in I}\sigma_{i,n}\right\}\sum_{j\in J}\mu_j\|L_j\|^2\right)<1. \tag{5.75}$$

Let $x_0\in\mathcal{H}$ and iterate

$$\begin{array}{l}\text{for } n=0,1,\ldots\\ \left\lfloor\begin{array}{l} y_{1,n}=x_n-\tau_n\Big(\sum_{i\in I}L_i^*v_{i,n}^*+\sum_{j\in J}L_j^*\big(\nabla h_j(L_jx_n-r_j)\big)\Big)\\ x_{n+1}=\operatorname{prox}_{\tau_n f}y_{1,n}\\ z_n=2x_{n+1}-x_n\\ \text{for every } i\in I\\ \left\lfloor\begin{array}{l} y_{2,i,n}=v_{i,n}^*+\sigma_{i,n}(L_iz_n-r_i)\\ v_{i,n+1}^*=\operatorname{prox}_{\sigma_{i,n}g_i^*}y_{2,i,n}.\end{array}\right.\end{array}\right.\end{array} \tag{5.76}$$

Proposition 5.36 [32] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.35. Then there exists a solution x to (5.13) such that $x_n\rightharpoonup x$.*

The following algorithm was first proposed in [2] in the case when $J=\varnothing$, which was extended in [27] to a block-coordinate and block-iterative asynchronous method. The following version, which explicitly exploits smooth functions, is proposed in [40].

Algorithm 5.37 Consider the setting of Problem 5.6 and let $(\gamma_n)_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $0<\inf_{n\in\mathbb{N}}\gamma_n\leqslant\sup_{n\in\mathbb{N}}\gamma_n<+\infty$. For every $k\in I\cup J$, let $v_{k,0}^*\in\mathcal{G}_k$, let $(\mu_{k,n})_{n\in\mathbb{N}}$ be a sequence in $]0,+\infty[$ such that $0<\inf_{n\in\mathbb{N}}\mu_{k,n}\leqslant\sup_{n\in\mathbb{N}}\mu_{k,n}<+\infty$, and let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $]0,2[$ such that $0<\inf_{n\in\mathbb{N}}\lambda_n\leqslant\sup_{n\in\mathbb{N}}\lambda_n<2$. Let $x_0\in\mathcal{H}$ and iterate

$$
\begin{array}{l}
\text{for } n = 0, 1, \ldots \\
\left\lfloor
\begin{array}{l}
l_n^* = \sum_{k \in I \cup J} L_k^* v_{k,n}^* \\
a_n = \operatorname{prox}_{\gamma_n f}(x_n - \gamma_n l_n^*) \\
a_n^* = \gamma_n^{-1}(x_n - a_n) - l_n^* \\
\text{for } i \in I \\
\left\lfloor
\begin{array}{l}
l_{i,n} = L_i x_n \\
b_{i,n} = r_i + \operatorname{prox}_{\mu_{i,n} g_i}\big(l_{i,n} + \mu_{i,n} v_{i,n}^* - r_i\big) \\
b_{i,n}^* = v_{i,n}^* + \mu_{i,n}^{-1}(l_{i,n} - b_{i,n}) \\
t_{i,n} = b_{i,n} - L_i a_n
\end{array}
\right. \\
\text{for } j \in J \\
\left\lfloor
\begin{array}{l}
l_{j,n} = L_j x_n \\
b_{j,n} = l_{j,n} - \mu_{j,n}(\nabla g_j(l_{j,n} - r_j) - v_{j,n}^*) \\
b_{j,n}^* = \nabla g_j(b_{j,n} - r_j) \\
t_{j,n} = b_{j,n} - L_j a_n
\end{array}
\right. \\
t_n^* = a_n^* + \sum_{k \in I \cup J} L_k^* b_{k,n}^* \\
\tau_n = \|t_n^*\|^2 + \sum_{k \in I \cup J} \|t_{k,n}\|^2 \\
\text{if } \tau_n = 0 \\
\left\lfloor
\begin{array}{l}
x = a_n \\
\text{for } k \in I \cup J \\
\lfloor \; v_k^* = v_{k,n}^* + \mu_{k,n}^{-1}(l_{k,n} - b_{k,n}) \\
\text{terminate.}
\end{array}
\right. \\
\text{if } \tau_n > 0 \\
\left\lfloor
\begin{array}{l}
\theta_n = \dfrac{\lambda_n}{\tau_n} \max\Big\{\langle x_n \mid t_n^* \rangle - \langle a_n \mid a_n^* \rangle + \sum_{k \in I \cup J}\big(\big\langle t_{k,n} \mid v_{k,n}^* \big\rangle - \big\langle b_{k,n} \mid b_{k,n}^* \big\rangle\big)\Big\} \\
x_{n+1} = x_n - \theta_n t_n^* \\
\text{for } k \in I \cup J \\
\lfloor \; v_{k,n+1}^* = v_{k,n}^* - \theta_n t_{k,n}.
\end{array}
\right.
\end{array}
\right.
\tag{5.77}
$$

Proposition 5.38 [40] *Let $(x_n)_{n\in\mathbb{N}}$ be a sequence generated by Algorithm 5.37. Then there exists a solution x to* (5.13) *such that $x_n \rightharpoonup x$.*

5.2.4 Application to inconsistent convex feasibility problems

As mentioned in the introduction, the convex feasibility formalism first proposed in [57] has enjoyed continued interest from the image recovery community [13, 22, 42, 51]. A structured formulation of this problem is the following.

Problem 5.39 Let $\mathcal{H}$ be a real Hilbert space, let K be a nonempty finite subset of $\mathbb{N}$, and let $(\mathcal{G}_k)_{k\in K}$ be a family of real Hilbert spaces. For every $k \in K$, suppose that $0 \neq L_k \in$

$\mathcal{B}(\mathcal{H}, \mathcal{G}_k)$, let $r_k \in \mathcal{G}_k$, and let C_k be a nonempty closed convex subset of $\mathcal{G}_k$. The goal is to

$$\text{find} \;\; x \in \mathcal{H} \;\; \text{such that} \;\; (\forall k \in K) \;\; L_k x - r_k \in C_k. \tag{5.78}$$

In applications, because of possible inaccuracies in *a priori* knowledge, unmodeled dynamics, or too aggressive confidence bounds on stochastic constraints, the above convex feasibility problem may turn out to be inconsistent [15, 21, 22, 37, 56], i.e.,

$$\bigcap_{k \in K} L_k^{-1}(r_k + C_k) = \varnothing. \tag{5.79}$$

In the presence of a possibly inconsistent feasibility problem, we propose the following variational formulation as a relaxation of (5.78).

Problem 5.40 Consider the setting of Problem 5.39. Let (I, J) be a partition of K such that, for every $i \in I$, $\phi_i \in \Gamma_0(\mathbb{R})$ is an even function that vanishes only at 0 and, for every $j \in J$, $\psi_j \colon \mathbb{R} \to \mathbb{R}$ is a differentiable convex even function that vanishes only at 0 with a Lipschitzian derivative. The problem is to

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \sum_{i \in I} \phi_i\big(d_{C_i}(L_i x - r_i)\big) + \sum_{j \in J} \psi_j\big(d_{C_j}(L_j x - r_j)\big). \tag{5.80}$$

Problem 5.40 unifies several formulations that have been proposed in the literature as surrogates to the possibly inconsistent Problem 5.39.

— If (5.78) happens to have solutions, they are the same as those of (5.80).

— Suppose that $J = \varnothing$ and $(\forall i \in I)\; \phi_i = \iota_{\{0\}}$. Then (5.80) reverts to (5.78).

— Let $(\omega_j)_{j \in J}$ be real numbers in $]0, 1]$ such that $\sum_{j \in J} \omega_j = 1$. Suppose that $I = \varnothing$ and $(\forall j \in J)\; \mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}$, $r_j = 0$, and $\psi_j = \omega_j |\cdot|^2/2$. Then we recover the least-squares formulation of [21], namely

$$\underset{x \in \mathcal{H}}{\text{minimize}} \;\; \frac{1}{2} \sum_{j \in J} \omega_j d_{C_j}^2(x). \tag{5.81}$$

— Let $(\omega_j)_{j \in J}$ be real numbers in $]0, 1]$ such that $\sum_{j \in J} \omega_j = 1$. Suppose that $I = \{0\}$, $\phi_0 = \iota_{\{0\}}$, and $(\forall j \in J)\; \mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}$, $r_j = 0$, and $\psi_j = \omega_j |\cdot|^2/2$. Then we recover the hard-constrained least-squares formulation of [25], namely

$$\underset{x \in C_0}{\text{minimize}} \;\; \frac{1}{2} \sum_{j \in J} \omega_j d_{C_j}^2(x). \tag{5.82}$$

For $J = \{1\}$, this framework reduces to the formulation

$$\underset{x \in C_0}{\text{minimize}} \;\; \frac{1}{2} d_{C_1}^2(x) \tag{5.83}$$

proposed in [37, 56].

— Let $(\omega_j)_{j\in J}$ be real numbers in $]0,1]$ such that $\sum_{j\in J}\omega_j = 1$. Suppose that $I=\{0\}$, $\phi_0 = \iota_{\{0\}}$, and (J_1, J_2) is a partition of J. Suppose further that $(\forall j\in J_1)$ $\mathcal{G}_j = \mathcal{H}$, $L_j = \mathrm{Id}\,$, $r_j = 0$, and $\psi_j = \omega_j|\cdot|^2/2$, and that $(\forall j\in J_2)$ $r_j = 0$ and $\psi_j = \omega_j|\cdot|^2/2$. Then we recover the formulation of [14], namely

$$\underset{x\in C_0}{\text{minimize}}\;\; \frac{1}{2}\sum_{j\in J_1}\omega_j d_{C_j}^2(x) + \frac{1}{2}\sum_{j\in J_2}\omega_j d_{C_j}^2(L_jx). \tag{5.84}$$

— Let $(\mathcal{H}_\ell)_{1\leqslant\ell\leqslant m}$ and $(\mathbb{K}_k)_{1\leqslant k\leqslant p}$ be real Hilbert spaces. For every $\ell\in\{1,\ldots,m\}$ and $k\in\{1,\ldots,p\}$, let D_ℓ be a nonempty closed convex subset of $\mathcal{H}_\ell$, let $\theta_\ell\in\Gamma_0(\mathbb{R})$ be even and vanishing only at 0, let E_k be a nonempty closed convex subset of $\mathbb{K}_k$, let $\rho_k\in\,]0,+\infty[$, and let $M_{k\ell}\in\mathcal{B}\,(\mathcal{H}_\ell,\mathbb{K}_k)$. Consider the multivariate convex feasibility problem

$$\text{find}\;\; x_1\in D_1,\ldots,x_m\in D_m \;\;\text{such that}\;\; \sum_{\ell=1}^m M_{1\ell}x_\ell\in E_1,\ldots,\sum_{\ell=1}^m M_{p\ell}x_\ell\in E_p. \tag{5.85}$$

In [12, Section 3.3], this problem was relaxed to

$$\underset{x_1\in\mathcal{H}_1,\ldots,x_m\in\mathcal{H}_m}{\text{minimize}}\;\; \sum_{\ell=1}^m\theta_\ell\big(d_{D_\ell}(x_\ell)\big) + \sum_{k=1}^p\frac{1}{2\rho_k}d_{E_k}^2\Bigg(\sum_{\ell=1}^m M_{k\ell}x_\ell\Bigg). \tag{5.86}$$

Now let

$$\begin{cases}\mathcal{H} = \bigoplus_{\ell=1}^m\mathcal{H}_\ell,\; \mathcal{G}_1 = \mathcal{H},\; \mathcal{G}_2 = \bigoplus_{k=1}^p\mathbb{K}_k\\ C_1 = D_1\times\cdots\times D_m,\; C_2 = E_1\times\cdots\times E_p\\ \phi_1\colon\mathbb{R}^m\to\,]-\infty,+\infty]\,\colon(\xi_1,\ldots,\xi_m)\mapsto\sum_{\ell=1}^m\theta_\ell(\xi_\ell)\\ \psi_2\colon\mathbb{R}^p\to\mathbb{R}\colon(\eta_1,\ldots,\eta_p)\mapsto\sum_{k=1}^p\eta_k^2/(2\rho_k)\\ L_1 = \mathrm{Id}\,,\; L_2\colon\mathcal{H}\to\mathcal{G}_2\colon(x_1,\ldots,x_m)\mapsto\big(\sum_{\ell=1}^m M_{1\ell}x_\ell,\ldots,\sum_{\ell=1}^m M_{p\ell}x_\ell\big).\end{cases} \tag{5.87}$$

Then $\phi_1\in\Gamma_0(\mathbb{R})$ vanishes only at zero and so does the smooth convex function ψ_2. Moreover, (5.85) reduces to

$$\text{find}\;\; x\in\mathcal{H}\;\;\text{such that}\;\; L_1x\in C_1\;\;\text{and}\;\; L_2x\in C_2, \tag{5.88}$$

which is an instance of (5.79), while (5.86) reduces to

$$\underset{x\in\mathcal{H}}{\text{minimize}}\;\; \phi_1\big(d_{C_1}(L_1x)\big)+\psi_2\big(d_{C_2}(L_2x)\big), \tag{5.89}$$

which is an instance of (5.80) with $I=\{1\}$ and $J=\{2\}$.

Let us note that Problem 5.40 corresponds to the special case of Problem 5.6 in which

$$f=0,\quad (\forall i\in I)\quad g_i=\phi_i\circ d_{C_i},\quad\text{and}\quad(\forall j\in J)\quad h_j=\psi_j\circ d_{C_j}. \tag{5.90}$$

To solve it with a fully proximal algorithm, one can invoke Lemma 5.9 for the nonsmooth terms, and Example 5.10 or Example 5.17 for the smooth ones.

5.2.5 Numerical illustrations

The objective of this section is to illustrate the viewpoint formulated in the Introduction, which suggests that it may be computationally advantageous to activate smooth functions proximally in certain instances.

We consider digital image restoration problems. All images have 128×128 pixels and therefore the underlying Hilbert space is $\mathcal{H} = \mathbb{R}^N$ ($N = 128^2$) equipped with the standard Euclidean norm $\|\cdot\|_2$. We have run many instances of each problem under various configurations, and the limited numerical results we present here are representative of the behavior one can expect. Note that the results are given in terms of iteration numbers as the algorithms compared in each experiment have similar execution time per iteration.

We shall use in particular Example 5.8. Note that in this context the linear operators model convolutions and they are representable by block-circulant matrices. The computation of Q in (5.21) is therefore straightforward via the fast Fourier transform [3].

5.2.5.1 Forward-backward versus Douglas-Rachford splitting

We consider a very basic instance of Problem 5.6 with only two functions. More specifically, we compare the numerical behavior of the forward-backward algorithms of Propositions 5.28 and 5.30 with that of the Douglas-Rachford algorithm of Proposition 5.32, which is a fully proximal method.

The original image is $\overline{x}$ and the degraded image is

$$y = H\overline{x} + w, \tag{5.91}$$

where H models a convolution with a uniform rectangular kernel of size 15×5 and w is a Gaussian white noise realization (see Fig. 5.1(a)–(b)). The blurred image-to-noise-ratio is 15.5 dB. Since each pixel value is known to be in $[0, 255]$, we use the hard constraint set $C = [0, 255]^N$. As is customary, the natural sparsity of $\overline{x}$ is promoted using the function $\|\cdot\|_1$. Altogether, the problem is to

$$\underset{x \in C}{\text{minimize}} \;\; \|x\|_1 + \frac{1}{2}\|Hx - y\|_2^2. \tag{5.92}$$

Now set $f = \|\cdot\|_1 + \iota_C$ and $g = \|H \cdot -y\|_2^2/2$, and let $\gamma \in \left]0, +\infty\right[$. Then $f \in \Gamma_0(\mathcal{H})$ has an explicit proximity operator $\text{prox}_{\gamma f} = \text{proj}_C \circ \text{soft}_\gamma$ [8, Propositions 24.12(ii) and 24.47], where soft_γ is the soft thresholder on $[-\gamma, \gamma]$. On the other hand, g is smooth and its proximity operator is provided by Example 5.8. We can therefore solve (5.92) with the forward-backward algorithms (5.71) and (5.72), as well as with the fully proximal Douglas-Rachford algorithm (5.73).

The algorithms are initialized at zero and implemented with parameters for which they seem to perform better, that is, $\gamma_n \equiv 1.9/\beta$ for (5.71), $\gamma = 1/\beta$ and $\alpha = 3$ for (5.72), and $\gamma = 30$ and $\lambda_n \equiv 1.9$ for (5.73). The results of Figs. 5.1–5.2 show a superior performance for the fully proximal algorithm (5.73). Note from Propositions 5.28 and 5.30 that the inertial forward-backward algorithm (5.72) is more demanding in terms of memory requirements than its standard counterpart (5.71), but it has a better theoretical upper bound on the rate of convergence of the objective function. However, as shown in Fig. 5.3, in this example, the actual behaviors of the methods in terms of the decrease of the objective function are quite comparable.

Remark 5.41 In the context of an inconsistent convex feasibility problem with two nonempty closed convex sets C_0 and C_1, the forward-backward algorithm (5.71) was applied to (5.83), which resulted in the alternating projection method $(\forall n \in \mathbb{N})$ $x_{n+1} = (\mathrm{proj}_{C_0} \circ \mathrm{proj}_{C_1})x_n$. This method was compared to the form of the Douglas-Rachford algorithm studied in [9] to solve (5.83). There too, the fully proximal Douglas-Rachford algorithm was reported to be faster than the forward-backward algorithm.

5.2.5.2 Primal-dual splitting

We consider a scenario similar to that of Section 5.2.5.1, but with a different original image $\overline{x}$ and two observations (see Fig. 5.4(a)–(c))

$$y_1 = H_1\overline{x} + w_1 \quad \text{and} \quad y_2 = H_2\overline{x} + w_2. \tag{5.93}$$

Here, H_1 and H_2 model convolution blurs with kernels of size 3×11 and of 7×5, respectively, and w_1 and w_2 are Gaussian white noise realizations. The blurred image-to-noise-ratios are 27.3 dB and 35.4 dB.

We use $C = [0, 255]^N$ as a hard constraint set. We assume that the diffraction of $\overline{x}$ has been partially observed over some frequency range K, possibly with measurement errors. To exploit this information we use the soft constraint penalty d_E^2, where ($\widehat{x}$ denotes the two-dimensional discrete Fourier transform) $E = \big\{x \in \mathbb{R}^N \;\big|\; \widehat{x}1_K = \widehat{\overline{x}}1_K\big\}$. The set K contains the frequencies in $\{0, \ldots, N/8 - 1\}^2$ as well as those resulting from the symmetry properties of the discrete Fourier transform. Finally, we use the total variation penalty to control oscillations in the restored image. This leads to the formulation

$$\underset{x \in C}{\text{minimize}}\; \frac{1}{2}d_E(x) + \frac{2}{5}h(Dx) + \frac{3}{4}\|H_1x - y_1\|^2 + \frac{3}{4}\|H_2x - y_2\|^2, \tag{5.94}$$

where $D\colon \mathbb{R}^N \to \mathbb{R}^N \times \mathbb{R}^N\colon x \mapsto (G_1x, G_2x)$, G_1 and G_2 being horizontal and vertical discrete difference operators, and where $(\forall y_1 \in \mathbb{R}^N)(\forall y_2 \in \mathbb{R}^N)$ $h(y_1, y_2) = \sum_{k=1}^{N} \phi(\|(\eta_{1,k}, \eta_{2,k})\|_2)$, ϕ being the Huber function (5.34) with parameter $\rho = 2$. We derive from (5.94) two versions of Problem 5.6.

Problem 5.42 In Problem 5.6, set $I = \{1\}$, $g_1 = 0.5d_C$, $L_1 = \mathrm{Id}$, $r_1 = 0$, $J = \{2, 3, 4\}$, $h_2 = 0.4h$, $L_2 = D$, $r_2 = 0$, $h_3 = 0.75\|\cdot\|_2^2$, $L_3 = H_1$, $r_3 = y_1$, $h_4 = 0.75\|\cdot\|_2^2$, $L_4 = H_2$, and $r_4 = y_2$.

Problem 5.43 (fully proximal) In Problem 5.6, set $I = \{1, 2, 3, 4\}$, $J = \varnothing$, $g_1 = 0.5d_C$, $L_1 = \mathrm{Id}$, $r_1 = 0$, $g_2 = 0.4h$, $L_2 = D$, $r_2 = 0$, $g_3 = 0.75\|H_1 \cdot -y_1\|_2^2$, $L_3 = \mathrm{Id}$, $r_3 = 0$, $g_4 = 0.75\|H_2 \cdot -y_2\|_2^2$, $L_4 = \mathrm{Id}$, and $r_4 = 0$.

We apply to these two problems Algorithms 5.33, 5.35, and 5.37 with $f = \iota_C$ and all initial vectors set to 0. The following parameters are used, where $\beta = \sqrt{\sum_{i\in I}\|L_i\|^2 + \sum_{j\in J}\mu_j\|L_j\|^2}$ (these parameters were found to optimize the performance of each algorithm) :

- Algorithm 5.33 : $\gamma_n \equiv 0.99/\beta$.
- Algorithm 5.35-S (with smooth terms for Problem 5.42) : $\sigma_{1,n} \equiv 8/(5\beta)$ and $\tau_n \equiv 8/(5\beta)$.
- Algorithm 5.35-P (fully proximal for Problem 5.43) : $\sigma_{1,n} \equiv 1/(2\beta)$, $\sigma_{2,n} \equiv 1/(2\beta)$, $\sigma_{3,n} \equiv 3/\beta$, $\sigma_{4,n} \equiv 3/\beta$, and $\tau_n \equiv 1/\beta$.
- Algorithm 5.37-S (with smooth terms for Problem 5.42) : $\gamma_n \equiv 0.4$, $\mu_{1,n} \equiv 1.0$, $\mu_{2,n} \equiv 2.49$, $\mu_{3,n} \equiv 0.65$, $\mu_{4,n} \equiv 0.65$, and $\lambda_n \equiv 1.99$.
- Algorithm 5.37-P (fully proximal for Problem 5.43) : $\gamma_n \equiv 0.25$, $\mu_{1,n} \equiv 1.0$, $\mu_{2,n} \equiv 1.5$, $\mu_{3,n} \equiv 1.0$, $\mu_{4,n} \equiv 1.0$, and $\lambda_n \equiv 1.99$.

The proximity operators and the gradients used in these experiments follow from [8, Example 24.28], Example 5.26 (with $M = 2$ and $\mathbb{K} = \{1, \ldots, N\}$, and using [8, Example 24.9] to get the proximity operator of ϕ), Example 5.14, and Example 5.8. Observe that $\|D\|^2 = 8$ and $\|H_1\|^2 = \|H_2\|^2 = 1$. On the other hand, the functions $0.4h$, $0.75\|\cdot\|_2^2$ and $0.75\|\cdot\|_2^2$ are differentiable with a Lipschitzian gradient, and their Lipschitz constants are respectively 0.4, 0.75, and 0.75. Thus, all the assumptions required by the algorithms are satisfied. The results shown in Figs. 5.5 and 5.6 illustrate the better numerical behavior of the fully proximal algorithms compared to their smooth versions.

5.3 Bibliographie

[1] F. Abboud, E. Chouzenoux, J.-C. Pesquet, J.-H. Chenot, and L. Laborelli, *Dual block-coordinate forward-backward algorithm with application to deconvolution and deinterlacing of video sequences,* J. Math. Imaging Vision, 59 (2017), pp. 415–431.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set,* SIAM J. Optim., 24 (2014), pp. 2076–2095.

[3] H. C. Andrews and B. R. Hunt, *Digital Image Restoration*. Prentice-Hall, Englewood Cliffs, NJ, 1977.

[4] A. Y. ARAVKIN, J. V. BURKE, AND G. PILLONETTO, *Sparse/robust estimation and Kalman smoothing with nonsmooth log-concave densities : modeling, computation, and theory,* J. Mach. Learn. Res., 14 (2013), pp. 2689–2728.

[5] T. ASPELMEIER, C. CHARITHA, AND D. R. LUKE, *Local linear convergence of the ADMM/Douglas-Rachford algorithms without strong convexity and application to statistical imaging,* SIAM J. Imaging Sci., 9 (2016), pp. 842–868.

[6] J.-F. AUJOL AND C. DOSSAL, *Stability of over-relaxations for the forward-backward algorithm, application to FISTA,* SIAM J. Optim., 25 (2015), pp. 2408–2433.

[7] H. H. BAUSCHKE AND J. M. BORWEIN, *On projection algorithm for solving convex feasibility problems,* SIAM Rev., 38 (1996), pp. 367–426.

[8] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces,* 2nd ed. Springer, New York, 2017.

[9] H. H. BAUSCHKE, P. L. COMBETTES, AND D. R. LUKE, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces,* J. Approx. Theory, 127 (2004), pp. 178–192.

[10] A. BECK AND M. TEBOULLE, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems,* SIAM J. Imaging Sci., 2 (2009), pp. 183–202.

[11] R. BELLMAN, R. E. KALABA, AND J. A. LOCKETT, *Numerical Inversion of the Laplace Transform : Applications to Biology, Economics Engineering, and Physics.* Elsevier, New York, 1966.

[12] L. M. BRICEÑO-ARIAS AND P. L. COMBETTES, *Convex variational formulation with smooth coupling for multicomponent signal decomposition and recovery,* Numer. Math. Theory Methods Appl., 2 (2009), pp. 485–508.

[13] C. L. BYRNE, *Iterative Optimization in Inverse Problems.* CRC Press, Boca Raton, FL, 2014.

[14] Y. CENSOR, T. ELFVING, N. KOPF, AND T. BORTFELD, *The multiple-sets split feasibility problem and its applications for inverse problems,* Inverse Problems, 21 (2005), pp. 2071–2084.

[15] Y. CENSOR AND M. ZAKNOON, *Algorithms and convergence results of projection methods for inconsistent feasibility problems : A review,* Pure Appl. Funct. Anal., to appear.

[16] L. CHAÂRI, J.-C. PESQUET, A. BENAZZA-BENYAHIA, AND PH. CIUCIU, *A wavelet-based regularized reconstruction algorithm for SENSE parallel MRI with applications to neuroimaging,* Med. Image Anal., 15 (2011), pp. 185–201.

[17] A. CHAMBOLLE AND C. DOSSAL, *On the convergence of the iterates of the 'Fast iterative shrinkage/thresholding algorithm',* J. Optim. Theory Appl., 166 (2015), pp. 968–982.

[18] A. CHAMBOLLE AND T. POCK, *A first-order primal-dual algorithm for convex problems with applications to imaging,* J. Math. Imaging Vision, 40 (2011), pp. 120–145.

[19] C. CHAUX, P. L. COMBETTES, J.-C. PESQUET, AND V. R. WAJS, *A variational formulation for frame-based inverse problems,* Inverse Problems, 23 (2007), pp. 1495–1518.

[20] E. CHOUZENOUX, A. JEZIERSKA, J.-C. PESQUET, AND H. TALBOT, *A convex approach for image restoration with exact Poisson-Gaussian likelihood,* SIAM J. Imaging Sci., 8 (2015), pp. 2662–2682.

[21] P. L. COMBETTES, *Inconsistent signal feasibility problems : Least-squares solutions in a product space,* IEEE Trans. Signal Process., 42 (1994), pp. 2955–2966.

[22] P. L. COMBETTES, *The convex feasibility problem in image recovery,* in : Advances in Imaging and Electron Physics (P. Hawkes, Ed.), vol. 95, pp. 155–270. Academic Press, New York, 1996.

[23] P. L. COMBETTES, *Solving monotone inclusions via compositions of nonexpansive averaged operators,* Optimization, 53 (2004), pp. 475–504.

[24] P. L. COMBETTES, *Perspective functions : Properties, constructions, and examples,* Set-Valued Var. Anal., 26 (2018), pp. 247–264.

[25] P. L. COMBETTES AND P. BONDON, *Hard-constrained inconsistent signal feasibility problems,* IEEE Trans. Signal Process., 47 (1999), pp. 2460–2468.

[26] P. L. COMBETTES, L. CONDAT, J.-C. PESQUET, AND B. C. VŨ, *A forward-backward view of some primal-dual optimization methods in image recovery,* Proc. IEEE Int. Conf. Image Process., pp. 4141–4145. Paris, France, Oct. 27–30, 2014.

[27] P. L. COMBETTES AND J. ECKSTEIN, *Asynchronous block-iterative primal-dual decomposition methods for monotone inclusions,* Math. Programming, 168 (2018), pp. 645–672.

[28] P. L. COMBETTES AND J.-C. PESQUET, *A Douglas-Rachford splitting approach to nonsmooth convex variational signal recovery,* IEEE Journal of Selected Topics in Signal Processing, 1 (2007), pp. 564–574.

[29] P. L. COMBETTES AND J.-C. PESQUET, *Proximal splitting methods in signal processing,* in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering,* (H. H. Bauschke et al., eds), pp. 185–212. Springer, New York, 2011.

[30] P. L. COMBETTES AND J.-C. PESQUET, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,* Set-Valued Var. Anal., 20 (2012), pp. 307–330.

[31] P. L. COMBETTES, S. SALZO, AND S. VILLA, *Consistent learning by composite proximal thresholding,* Math. Program., B167 (2018), pp. 99–127.

[32] P. L. COMBETTES AND B. C. VŨ, *Variable metric forward-backward splitting with applications to monotone inclusions in duality,* Optimization, 63 (2014), pp. 1289–1318.

[33] P. L. COMBETTES AND V. R. WAJS, *Signal recovery by proximal forward-backward splitting,* Multiscale Model. Simul., 4 (2005), pp. 1168–1200.

[34] L. CONDAT, *A primal-dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms,* J. Optim. Theory Appl., 158 (2013), pp. 460–479.

[35] I. DAUBECHIES, M. DEFRISE, AND C. DE MOL, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,* Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457.

[36] C. DE MOL AND M. DEFRISE, *A note on wavelet-based inversion algorithms,* Contemp. Math., 313 (2002), pp. 85–96.

[37] M. GOLDBURG AND R. J. MARKS II, *Signal synthesis in the presence of an inconsistent set of constraints,* IEEE Trans. Circuits Syst., 32 (1985), pp. 647–663.

[38] B. HE AND X. YUAN, *Convergence analysis of primal-dual algorithms for a saddle-point problem : From contraction perspective,* SIAM J. Imaging Sci., 5 (2012), pp. 119–149.

[39] M. HINTERMÜLLER AND G. STADLER, *An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration,* SIAM J. Sci. Comput., 28 (2006), pp. 1–23.

[40] P. R. JOHNSTONE AND J. ECKSTEIN, *Projective splitting with forward steps : Asynchronous and block-iterative operator splitting,* arxiv, 2018. https://arxiv.org/pdf/1803.07043.pdf

[41] S. L. KEELING, *Total variation based convex filters for medical imaging,* Appl. Math. Comput., 139 (2003), pp. 101–119.

[42] Y. LIU, Z. LIANG, AND J. MA, *Total variation-Stokes strategy for sparse-view X-ray CT image reconstruction,* IEEE Trans. Med. Imaging, 33 (2014), pp. 749–763.

[43] C. LOUCHET AND L. MOISAN, *Posterior expectation of the total variation model : Properties and experiments,* SIAM J. Imaging Sci., 6 (2013), pp. 2640–2684.

[44] B. MARTINET, *Détermination approchée d'un point fixe d'une application pseudo-contractante. Cas de l'application prox,* C. R. Acad. Sci. Paris, A274 (1972), pp. 163–165.

[45] J. J. MOREAU, *Fonctions convexes duales et points proximaux dans un espace hilbertien,* C. R. Acad. Sci. Paris, A255 (1962), pp. 2897–2899.

[46] D. O'CONNOR AND L. VANDENBERGHE, *Primal-dual decomposition by operator splitting and applications to image deblurring,* SIAM J. Imaging Sci., 7 (2014), pp. 1724–1754.

[47] N. PAPADAKIS, G. PEYRÉ, AND E. OUDET, *Optimal transport with proximal splitting,* SIAM J. Imaging Sci., 7 (2014), pp. 212–238.

[48] H. RAGUET AND L. LANDRIEU, *Preconditioning of a generalized forward-backward splitting and application to optimization on graphs,* SIAM J. Imaging Sci., 8 (2015), pp. 2706–2739.

[49] C. SCHNÖRR, *Unique reconstruction of piecewise-smooth images by minimizing strictly convex nonquadratic functionals,* J. Math. Imaging Vision, 4 (1994), pp. 189–198.

[50] I. STEINWART, *Consistency of support vector machines and other regularized kernel classifiers,* IEEE Trans. Inform. Theory, 51 (2005), pp. 128–142.

[51] M. TOFIGHI, O. YORULMAZ, K. KÖSE, D. C. YILDIRIM, R. ÇETIN-ATALAY, AND A. E. ÇETIN, *Phase and TV based convex sets for blind deconvolution of microscopic images,* IEEE J. Selected Topics Signal Process., 10 (2016), pp. 81–91.

[52] R. S. VARGA, *Matrix Iterative Analysis,* 2nd edition. Springer-Verlag, New York, 2000.

[53] B. C. VŨ, *A splitting algorithm for dual monotone inclusions involving cocoercive operators,* Adv. Comput. Math., 38 (2013), pp. 667–681.

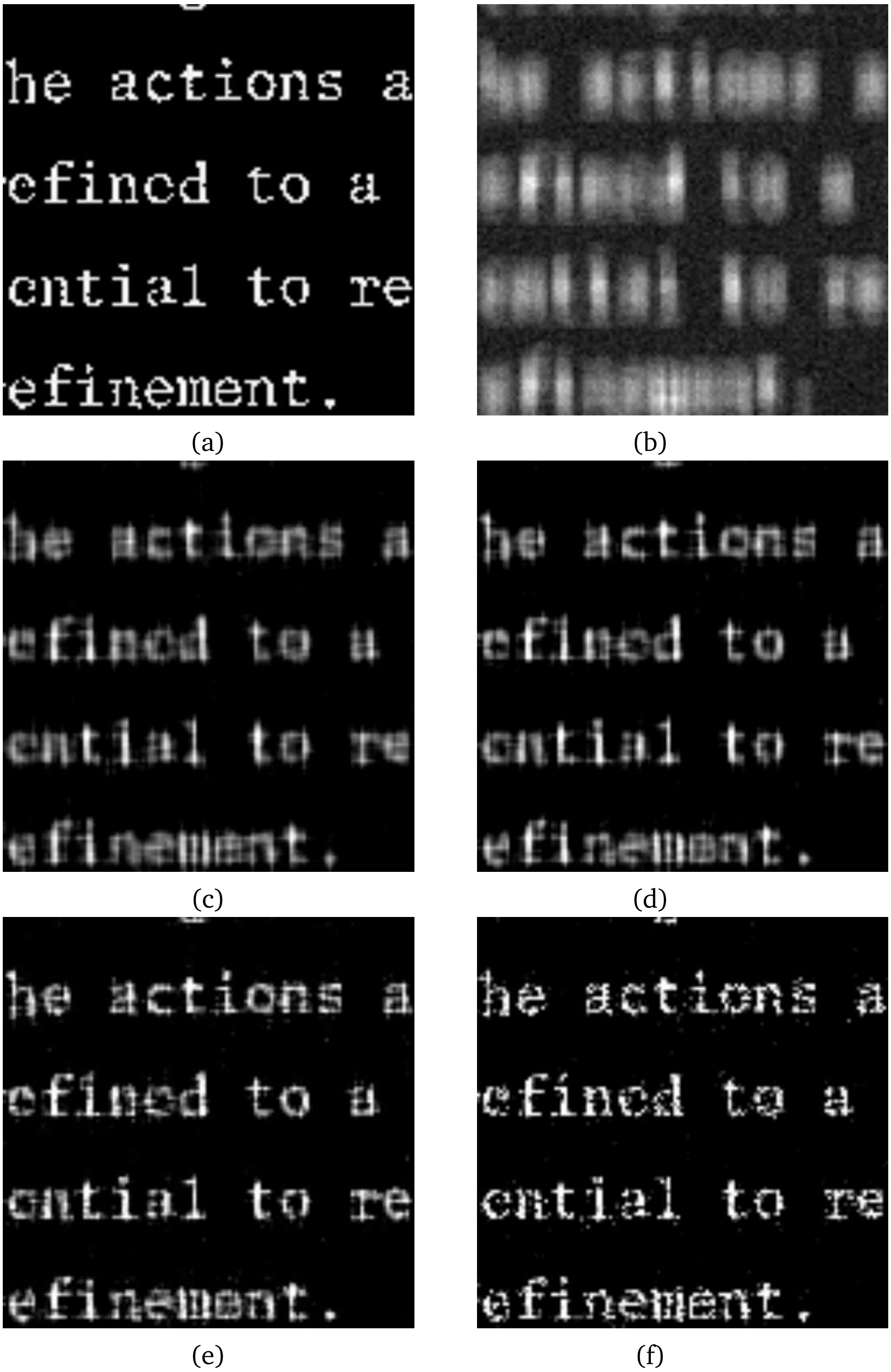
[54] B. C. VŨ, *A variable metric extension of the forward-backward-forward algorithm for monotone operators,* Numer. Funct. Anal. Optim., 34 (2013), pp. 1–16.

[55] D. C. YOULA, *Generalized image restoration by the method of alternating orthogonal projections,* IEEE Trans. Circuits Syst., 25 (1978), pp. 694–702.

[56] D. C. YOULA AND V. VELASCO, *Extensions of a result on the synthesis of signals in the presence of inconsistent constraints,* IEEE Trans. Circuits Syst., 33 (1986), pp. 465–468.

[57] D. C. YOULA AND H. WEBB, *Image restoration by the method of convex projections : Part 1 – theory,* IEEE Trans. Med. Imaging, 1 (1982), pp. 81–94.

[58] T. ZHANG, *Statistical behavior and consistency of classification methods based on convex risk minimization,* Ann. Stat., 32 (2004), pp. 56–85.



(a) (b) (c) (d) (e) (f)

FIGURE 5.1 – (a) Original image $\overline{x}$. (b) Degraded image y. (c) Image restored by the forward-backward algorithm (5.71) after 50 iterations. (d) Image restored by the inertial forward-backward algorithm (5.72) after 50 iterations. (e) Image restored by the Douglas-Rachford algorithm (5.73) after 50 iterations. (f) Restored image (all algorithms yield visually equivalent images).

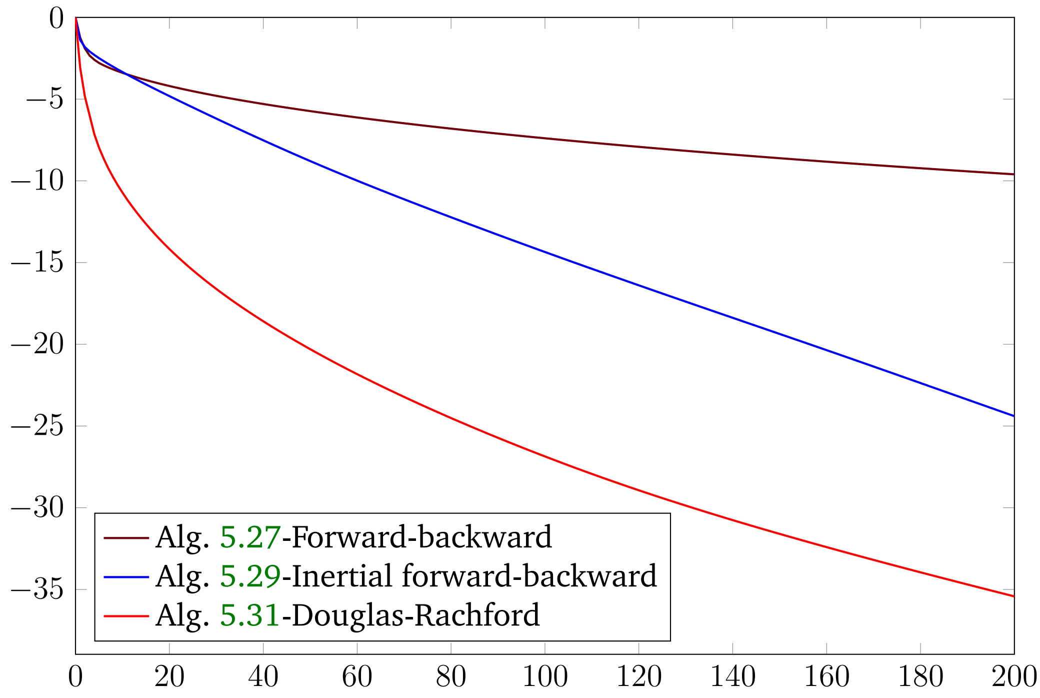


FIGURE 5.2 – Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number.

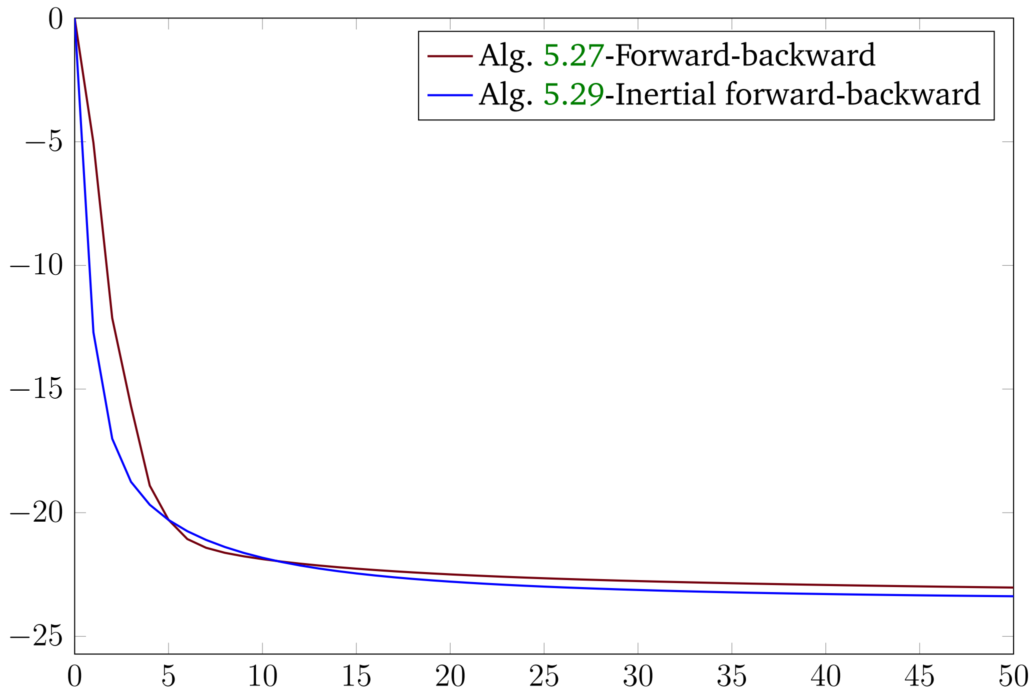


FIGURE 5.3 – Normalized objective function of (5.92) in dB versus iteration number.

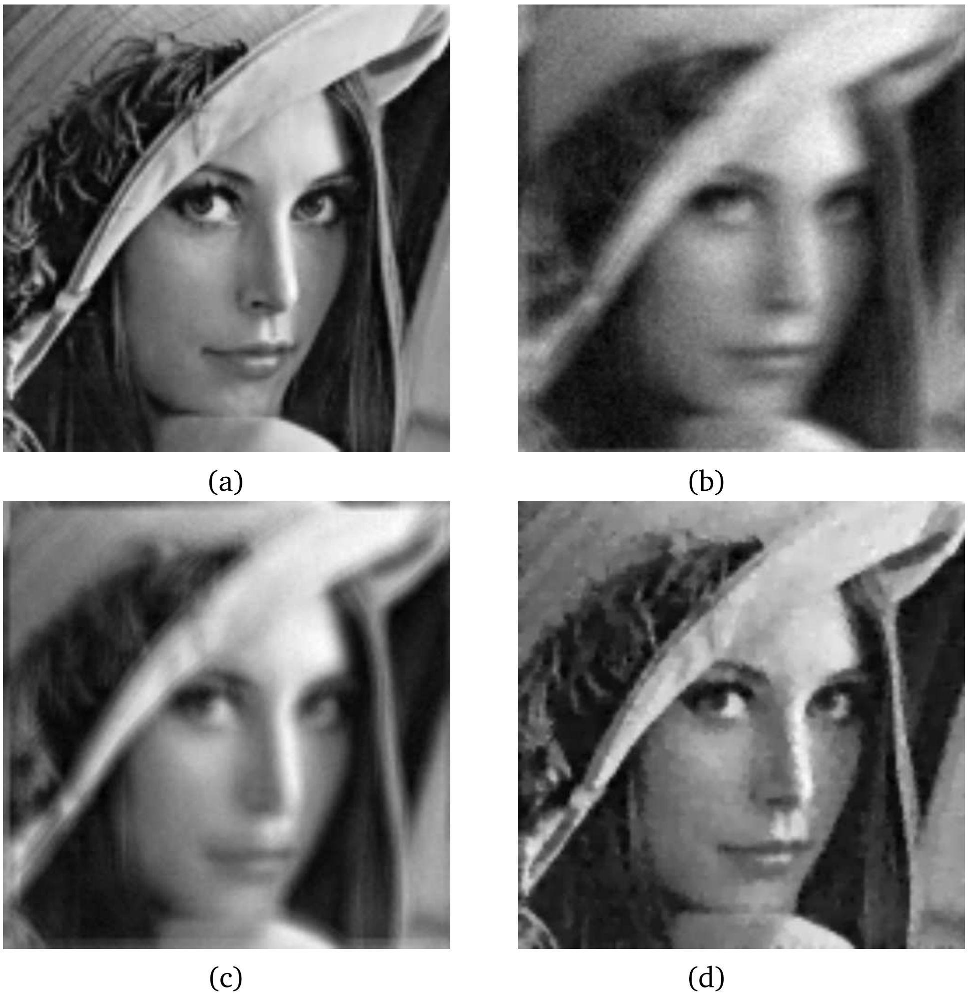


FIGURE 5.4 – (a) Original image $\overline{x}$. (b) Degraded image y_1. (c) Degraded image y_2. (d) Restored image (all algorithms yield visually equivalent images).

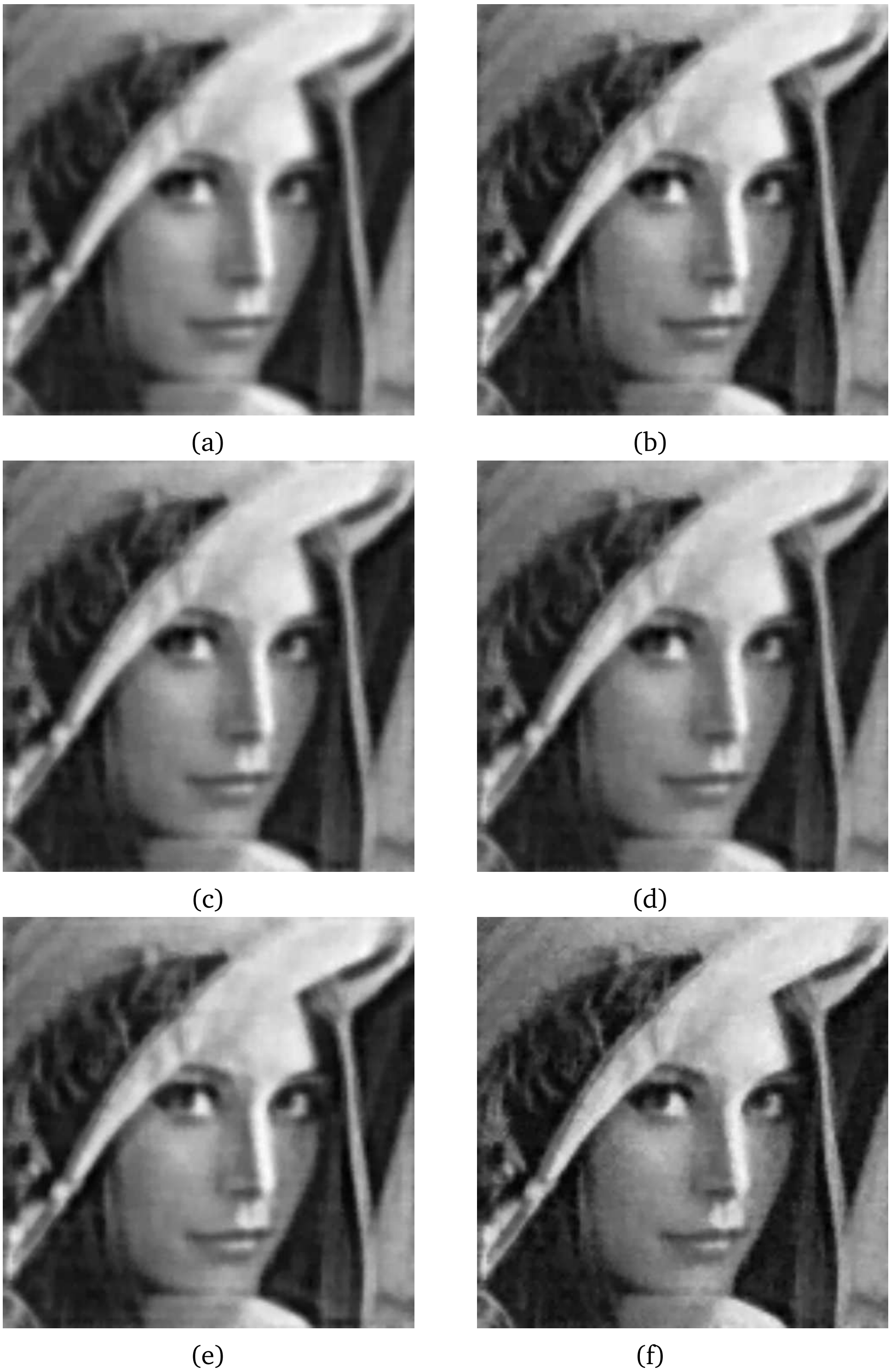


FIGURE 5.5 – (a) Image restored by Pb. 5.42/Alg. 5.33-S after 20 iterations. (b) Image restored by Pb. 5.43/Alg. 5.33-P after 20 iterations. (c) Image restored by Pb. 5.42/Alg. 5.35-S after 20 iterations. (d) Image restored by Pb. 5.43/Alg. 5.35-P after 20 iterations. (e) Image restored by Pb. 5.42/Alg. 5.37-S after 20 iterations. (f) Image restored by Pb. 5.43/Alg. 5.37-P after 20 iterations.

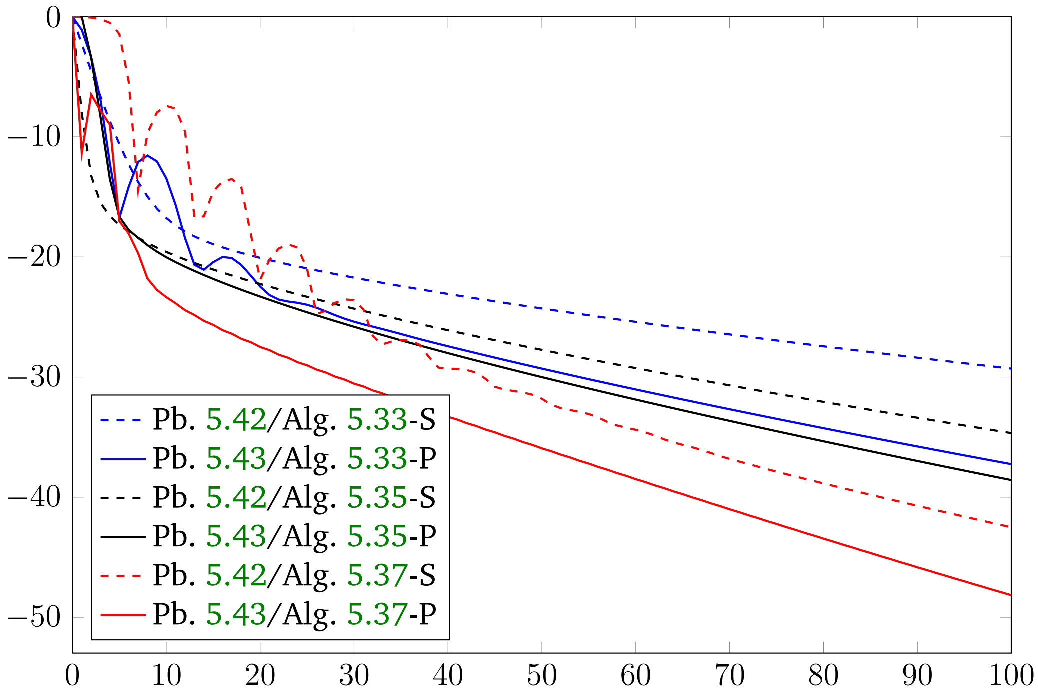


FIGURE 5.6 – Normalized distance in dB to the asymptotic image produced by each algorithm versus iteration number.

Chapitre 6

Conclusion et perspectives

6.1 Conclusion

Dans cette thèse, nous avons proposé un cadre général et souple pour l'étude des méthodes de point fixe qui englobe diverses stratégies de mise en œuvre. Ces méthodes peuvent être multicouches, avec mémoire, ou à métrique variable. Elles tolèrent de plus des erreurs et permettent des relaxations. Outre le développement de nouveaux algorithmes, cette étude nous a permis de rassembler dans un cadre unique et simplifé des résultats très divers concernant l'éclatement d'opérateurs monotones. Parmi les nouveaux algorithmes développés, nous avons étudié la première méthode pour résoudre des problèmes multicouches de point fixe à métrique variable, et l'avons appliquée à des problèmes d'inclusions monotones composites. En dernier lieu, nous avons proposé l'utilisation d'opérateurs proximaux pour des fonctions lisses alors qu'il est d'usage d'utiliser leur gradient. Nous avons aussi présenté une nouvelle formulation pour les problèmes incompatibles où, justement, les fonctions lisses peuvent être utilisées. Des simulations ont été fournies pour illustrer l'efficacité des algorithmes implicites par rapport aux algorithmes explicites.

6.2 Perspectives

- Dans le Chapitre 3, nous avons émis plusieurs conjectures : peut-on construire des matrices superconcentrantes d'ordre infini ? Est-il possible de montrer la convergence d'un algorithme explicite-implicite inertiel à au moins 3 pas de mémoire tout en conservant un taux de $O(1/n^2)$ sur la fonction objectif ? Il serait intéressant à la fois d'un point de vue théorique et numérique d'y répondre.
- La méthode d'Haugazeau généralisée permet d'obtenir un principe de convergence

forte pour des suites qui convergent faiblement [8]. Peut-on appliquer ce principe pour obtenir la convergence forte dans le cas des algorithmes avec mémoire généraux ou des algorithmes à métrique variable ?

- Pour un algorithme inertiel de la forme

$$
\begin{array}{l}
\text{pour } n = 0, 1, \ldots \\
\left\lfloor \begin{array}{l} \overline{x}_n = x_n + \eta_n(x_n - x_{n-1}) \\ x_{n+1} = T\overline{x}_n, \end{array} \right.
\end{array}
\tag{6.1}
$$

  la suite $(\overline{x}_n)_{n\in\mathbb{N}}$ vérifie

$$
(\forall n \in \mathbb{N}) \quad \overline{x}_{n+1} = T\overline{x}_n + \eta_{n+1}(T\overline{x}_n - T\overline{x}_{n-1}),
\tag{6.2}
$$

  ce qui peut être considéré comme une forme de sur-relaxation. Maintenant, prenons spécifiquement le cas de l'algorithme explicite-implicite. On sait qu'en diminuant le paramètre proximal $(\gamma_n)_{n\in\mathbb{N}}$ jusqu'à β, où β est la constante de cocoercivité de l'un des deux opérateurs (ce qu'on est obligé de faire dans les méthodes inertielles type FISTA pour assurer la convergence), la sur-relaxation $(\lambda_n)_{n\in\mathbb{N}}$ peut être augmentée jusqu'au seuil $3/2$, comme rappelé dans le Chapitre 4. Si on appelle $(\rho_n)_{n\in\mathbb{N}}$ la suite telle que $\lambda_n = 1 + \rho_n$, alors on a

$$
(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n(Tx_n - x_n) = Tx_n + \rho_n(Tx_n - x_n).
\tag{6.3}
$$

  On retrouve ainsi un schéma proche de l'inertiel, mais avec $(\rho_n)_{n\in\mathbb{N}}$ bornée par $0,5$, alors que la suite $(\eta_n)_{n\in\mathbb{N}}$ est habituellement bornée par 1. Ainsi il y a une différence entre ces deux relaxations qu'il serait utile de comprendre.

- Dans [14], des versions stochastiques de méthodes de point fixe multicouches ont été développées et appliquées à de nombreux algorithmes pour les inclusions monotones. Peut-on appliquer ce formalisme à nos méthodes avec mémoire ?

- Récemment les distances de Bregman ont été utilisées en éclatement d'opérateurs [6, 12, 16]. Ces méthodes permettent d'appliquer les méthodes classiques d'optimisation dans des cadres plus larges. L'extension de nos résultats à l'aide de ces outils serait intéressante à développer.

- Nous nous sommes principalement concentré sur les problèmes de décomposition de type $A + B$, où A et B sont deux opérateurs maximalement monotones, pour ne pas alourdir les exemples. Cependant, de nombreux problèmes récents font intervenir un opérateur linéaire borné, comme dans le problème composite suivant.

  Problème 6.1 Soient $A\colon \mathcal{H} \to 2^{\mathcal{H}}$ et $B\colon \mathcal{G} \to 2^{\mathcal{G}}$ deux opérateurs maximalement monotones et soit $L \in \mathcal{B}(\mathcal{H}, \mathcal{G})$. Supposons que $\operatorname{zer}(A + L^*BL) \neq \varnothing$. L'objectif est de

$$
\text{trouver } x \in \mathcal{H} \text{ tel que } 0 \in Ax + L^*BLx.
\tag{6.4}
$$

Nous disposons depuis une dizaine d'années de méthodes pour résoudre ce problème sans y imposer de contraintes supplémentaires [1, 2, 10, 11, 13, 15]. L'application de nos méthodes avec mémoire à la résolution de ces problèmes composites pourrait avoir un effet très positif sur la convergence.

- Les processus continus liés aux algorithmes inertiels ont été étudiés dans [3, 4, 5, 9]. Il serait intéressant d'interpréter nos résultats dans le cadre continu. La dynamique continue

$$(\forall t \in]0, +\infty[) \quad x(t) + \dot{x}(t) = Tx(t) \tag{6.5}$$

donne, après discrétisation,

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = x_n + \lambda_n (T(x_n) - x_n). \tag{6.6}$$

Dans notre cas, nous pouvons regarder la dynamique suivante

$$(\forall t \in]0, +\infty[) \quad y(t) + \dot{y}(t) = T\left(t, \int_0^t y(s)\mu_t d(s)\right) \tag{6.7}$$

et sa discrétisation

$$(\forall n \in \mathbb{N}) \quad y_{n+1} = y_n + \lambda_n (T_n(\overline{y}_n) - y_n), \tag{6.8}$$

qui est d'ailleurs differente de notre processus

$$(\forall n \in \mathbb{N}) \quad x_{n+1} = \overline{x}_n + \lambda_n (T_n \overline{x}_n - \overline{x}_n). \tag{6.9}$$

L'étude complète de (6.7) peut fournir de nombreux résultats nouveaux et permettre d'affiner le choix des matrices.

Paris, le 13 Décembre 2018

6.3 Bibliographie

[1] M. A. Alghamdi, A. Alotaibi, P. L. Combettes, and N. Shahzad, *A primal-dual method of partial inverses for composite inclusions,* Optim. Lett., 8 (2014), pp. 2271–2284.

[2] A. Alotaibi, P. L. Combettes, and N. Shahzad, *Solving coupled composite monotone inclusions by successive Fejér approximations of their Kuhn-Tucker set,* SIAM J. Optim., 24 (2014), pp. 2076–2095.

[3] H. Attouch and A. Cabot, *Convergence of damped inertial dynamics governed by regularized maximally monotone operators,* J. Differential Equations, 264 (2018), pp. 7138–7182.

[4] H. Attouch, Z. Chbani, J. Peypouquet, and P. Redont, *Fast convergence of inertial dynamics and algorithms with asymptotic vanishing viscosity,* Math. Program., 168 (2018), pp. 123–175.

[5] J.-F. Aujol, C. Dossal, and A. Rondepierre, *Optimal convergence rates for Nesterov acceleratiion*, arxiv, 2018. https://arxiv.org/pdf/1805.05719.pdf

[6] H. H. Bauschke, J. Bolte, and M. Teboulle, *A descent lemma beyond Lipschitz gradient continuity : first-order methods revisited and applications, Math. Oper. Res.,* 42 (2018), pp. 330–348.

[7] H. H. Bauschke, J. M. Borwein, and P. L. Combettes, *Bregman monotone optimization algorithms,* SIAM J. Control Optim., 42 (2003), pp. 596–636.

[8] H. H. Bauschke and P. L. Combettes, *A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces,* Math. Oper. Res., 26 (2001), pp. 248–264.

[9] R. I. Boţ and E. R. Csetnek, *Second order forward-backward dynamical systems for monotone inclusion problems,* SIAM J. Control Optim., 54 (2016), pp. 1423–1443.

[10] L. M. Briceño-Arias and P. L. Combettes, *A monotone+skew splitting model for composite monotone inclusions in duality,* SIAM J. Optim., 21 (2011), pp. 1230–1250.

[11] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging,* J. Math. Imaging Vis., 40 (2011), pp. 120–145.

[12] P. L. Combettes and Q. V. Nguyen, Solving composite monotone inclusions in reflexive Banach spaces by constructing best Bregman approximations from their Kuhn-Tucker set, J. Convex Anal., 23 (2016), pp. 481–510.

[13] P. L. Combettes and J.-C. Pesquet, *Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators,* Set-Valued Var. Anal., 20 (2012), pp. 307–330.

[14] P. L. Combettes and J.-C. Pesquet, *Stochastic quasi-Fejér block-coordinate fixed point iterations with random sweeping,* SIAM J. Optim., 25 (2015), pp. 1221–1248.

[15] P. L. Combettes and B. C. Vũ, *Variable metric quasi-Fejér monotonicity,* Nonlinear Anal., 78 (2013), pp. 17–31.

[16] Q. V. Nguyen, *Variable quasi-Bregman monotone sequences,* Numer. Algorithms, 73 (2016), pp. 1107–1130.