Convergence Rates with Inexact Non-expansive Operators

Jingwei Liang*, Jalal Fadili*, Gabriel Peyré†

Abstract

In this paper, we present a convergence rate analysis for the inexact Krasnosel'skiĭ–Mann iteration built from nonexpansive operators. Our results include two main parts: we first establish global pointwise and ergodic iteration–complexity bounds, and then, under a metric subregularity assumption, we establish local linear convergence for the distance of the iterates to the set of fixed points. The obtained iteration–complexity result can be applied to analyze the convergence rate of various monotone operator splitting methods in the literature, including the Forward–Backward, the Generalized Forward–Backward, Douglas–Rachford, ADMM and Primal–Dual splitting methods. For these methods, we also develop easily verifiable termination criteria for finding an approximate solution, which can be seen as a generalization of the termination criterion for the classical gradient descent method. We finally develop a parallel analysis for the non-stationary Krasnosel'skiĭ–Mann iteration. The usefulness of our results is illustrated by applying them to a large class of structured monotone inclusion and convex optimization problems. Experiments on some large scale inverse problems in signal and image processing problems are shown.

Key words. Monotone inclusion, Non-expansive operator, Convergence rates, Operator splitting, Convex optimization, Inverse problems.

1 Introduction

1.1 Monotone inclusion and operator splitting

In various fields of science and engineering, such as signal/image processing and machine learning, many problems can be cast as solving a structured monotone inclusion problem. A prototypical example that has attracted a wave of interest recently, see *e.g.* [68, 20, 57], takes the form

$$\text{Find } x \in \mathcal{H} \text{ such that } 0 \in Bx + \sum_{i=1}^{n} L_i^* \circ A_i \circ L_i x, \tag{1.1}$$

where B is cocoercive, A_i is a set–valued maximal monotone operator acting on a real Hilbert space $\mathcal{G}_i$ with scalar product and associated norm, and L_i is a bounded linear operator from $\mathcal{H}$ to $\mathcal{G}_i$. Even more complex problems (*e.g.* with parallel sums) will be discussed in detail in Section 5.

The first operator splitting method has been developed from the 70's to solve structured monotone inclusion problems. Since then, the class of splitting methods have been regularly enriched with increasingly sophisticated algorithms as the structure of problems to handle becomes more complex. Splitting methods are iterative algorithms which may evaluate (possibly

*Jingwei Liang, Jalal Fadili
GREYC, CNRS–ENSICAEN–Université de Caen, E-mail: {jingwei.liang,jalal.fadili}@ensiacen.fr
†Gabriel Peyré
CNRS, CEREMADE, Université Paris–Dauphine, E-mail: peyre@ceremade.dauphine.f

approximately) the individual operators, their resolvents, the linear operators, all separately at various points in the course of iteration, but never the resolvents of sums nor of composition by a linear operator. Popular splitting algorithms to solve special instances of (1.1) include the Forward–Backward splitting method (FBS) [47, 54], the Douglas–Rachford splitting method (DRS) [27, 44] and Peaceman–Rachford splitting method (PRS) [55], ADMM [29, 30, 31, 32] which is DRS applied to the dual [28]. Other splitting methods were designed to solve (1.1) or even more complex problems (*e.g.* with parallel sums), see for instance [15, 66, 62, 8, 14, 57, 20, 23, 68].

Many splitting algorithms (including FBS, DRS, ADMM, and many others) can be cast as the Krasnosel'skiĭ–Mann fixed point iteration [38, 45] perhaps in its inexact form to handle errors,

$$z^{k+1} = z^k + \lambda_k\big(Tz^k + \varepsilon^k - z^k\big)$$

where T is a non-expansive operator acting on an appropriate Hilbert space, $(\lambda_k)_{k\in\mathbb{N}} \in]0,1[$, and ε^k is the error when computing Tz^k. The sequence z^k is built in way that it can be easily related to the original sequence x^k.

1.2 Contributions

In this paper, for the inexact Krasnosel'skiĭ–Mann iteration built from non-expansive operators with absolutely summable errors, global pointwise and ergodic iteration–complexity bounds are presented. Then under a metric subregularity [26] assumption, we establish local linear convergence of the iteration in the sense that the distance of the iterate to the set of fixed points converges linearly. Of course, when the fixed point is unique, the sequence itself converges linearly to it.

Our result can be applied to analyze the convergence behaviour of the iterates generated by a variety of monotone operator splitting methods. One crucial property of these methods is that they have an equivalent fixed point formulation whose corresponding fixed point operator is non-expansive[1]. Such methods include the FBS, Generalized Forward–Backward (GFB) [57], DRS, ADMM, and several Primal–Dual splitting (PDS) methods [68, 17]. In particular, for the GFB method developed by two of the co-authors, which adresses the case when L_i's in (1.1) equal to identity, we demonstrate that $O(1/\epsilon)$ iterations are needed to find a pair $\big((u_i)_i, g\big)$ with the termination criterion $\|g + B(\sum_i \omega_i u_i)\|^2 \leq \epsilon$, where $g \in \sum_i A_i(u_i)$. This termination criterion can be viewed as a generalization of the classical one based on the norm of the gradient for the gradient descent method [52]. The iteration–complexity improves to $O(1/\sqrt{\epsilon})$ in ergodic sense for the same termination criterion. Similar interpretation is also provided for the DRS/ADMM and PDS considered in Section 5.

We finally study the iteration–complexity of the non-stationary version of the Krasnosel'skiĭ–Mann iteration, by appropriatly absorbing the non-stationarity into an additional error term. Under reasonable conditions, we show that the iteration–complexity bounds developed above remain valid. We illustrate this on the GFB method.

1.3 Related work

1.3.1 Global iteration–complexity bounds

Relation with [33] In (1.1), if $B = 0$, $n = 2$ and $L_i = \mathrm{Id}$, $i = 1, 2$, then the problem can be solved by the DRS algorithm described in Algorithm 2. The iteration–complexity of the exact

[1] In fact, in many cases it is even α–averaged, see Definition 2.1.

DRS is studied in [33], under the assumption that A_2 is single–valued, an error term is defined,

$$e^k = z^k - J_{\gamma A_1}(\mathrm{Id} - \gamma A_2)z^k,$$

where z^k is the iterate and $J_{\gamma A_1} = (\mathrm{Id} + \gamma A_1)^{-1}$ is the resolvent of A_1. By relying on firm non-expansiveness of the resolvent [2], the authors show that $\|e^k\|$ converges to 0 at the rate of $O(1/\sqrt{k})$. In fact, it can be easily verified that

$$e^k = z^k - \frac{1}{2}\big((2J_{\gamma A_1} - \mathrm{Id})(2J_{\gamma A_2} - \mathrm{Id}) + \mathrm{Id}\big)z^k,$$

without assuming A_2 to be single–valued. The operator $\frac{1}{2}((2J_{\gamma A_1} - \mathrm{Id})(2J_{\gamma A_2} - \mathrm{Id}) + \mathrm{Id})$ is firmly non-expansive, hence (1/2)–averaged non-expansive. Our results in Section 3 goes much beyond this work by considering a more general iterative scheme with an operator that is only non-expansive and that may be evaluated approximately.

Relation with [22] In [22], the authors consider the exact Krasnosel'skiĭ–Mann iteration and showed that $\|z^k - z^{k-1}\| = O(1/\sqrt{k})$. Our work differs from [22] in 3 aspects: 1) we consider the inexact scheme; 2) we provide a shaper monotonicity property compared to [22, Proposition 11]; 3) we establish pointwise and ergodic iteration–complexity bounds as well as local linear convergence analysis.

Relation with [63] Based on the enlargement of maximal monotone operators, in [63], a hybrid proximal extragradient method (HPE) is introduced to solve monotone inclusion problem of the form

$$\text{Find } x \in \mathcal{H} \ \text{ such that } \ 0 \in Ax.$$

The HPE framework encompasses some splitting algorithms of the literature, see [49]. Convergence of HPE is established in [63] and in [9] for its inexact version. The pointwise and ergodic iteration–complexities of the exact HPE on a similar error criterion as in our work were established in [49]. Some of the splitting methods we consider in Section 5 are also covered by HPE, and thus our iteration complexity bounds coincide with those of HPE, but only for the ergodic case. In the pointwise case, our bound is uniform while theirs is not.

1.3.2 Local linear convergence

Relation with [41, 43] In [41], local lineare convergence of the distance to the set of zeros of a maximal monotone operator using the exact proximal point algorithm (PPA [46, 59]) is established by assuming metric subregularity of the operator. Local convergence rate analysis of PPA under a higher–order extension of metric subregularity, namely metric q–subregularity $q \in]0, 1]$, is conducted in [43]. In our work, metric subregularity is assumed on $\mathrm{Id} - T$, where T is the fixed point operator, *i.e.* the resolvent, rather than the maximal monotone operator in the case of PPA. Relation between metric subregularity of these operators is intricate in general and is beyond the scope of this paper, though we provide an instructive discussion for a simple case at the end of Section 4. Note also that the work of [41, 43] considers PPA only in its classical form, *i.e.* without errors nor relaxation.

Local linear rate for feasibility problems Based on strong regularity, [42] proved local linear convergence of the Method of Alternating Projections (MAP) in the non-convex setting, where the sets are closed and one of which is suitably regular. The linear rate is associated with a modulus of regularity. This is refined later in [3]. In [35], the authors develop local linear convergence results for the MAP and DRS to solve non-convex feasibility problems. Their analysis relies on a local version of firm non-expansiveness together with a coercivity condition. It turns out that this coercivity condition holds for mapping T for which the fixed points are isolated and $\mathrm{Id}-T$ is metrically regular [36, Lemma 25]. The linear rate they establish, however, imposes a bound on the metric regularity modulus.

Other local linear rates with DRS For the case of a sphere intersecting a line, or more generally a proper affine subset, typically in $\mathbb{R}^2$, [4] establishes local linear convergence of DRS. Local linear convergence of (the relaxed) DRS to solve the affinely constraint ℓ_1–minimization problem (basis pursuit) is shown in [24]. For this particular instance, their interiority assumption can be related to metric subregularity of the DRS fixed point operator. Given the level of details this relation requires, we defer it (in an even more general setting) to a forthcoming paper, hence we do not dwell on it further.

1.4 Paper organization

The organization of the paper is as follows. In Section 2 we recall some preliminary results on monotone operator theory and define the product space. Global iteration–complexity bounds and local convergence rate of the inexact Krasnosel'skiĭ–Mann iteration are established in Section 3 and Section 4 respectively. In Section 5, illustrative examples of existing monotone operator splitting methods to which our iteration–complexity result can be applied are described. Extension to the non-stationary fixed point iterations is discussed in Section 6. Some numerical experiments are shown in Section 7. The proofs are collected in the appendix.

2 Preliminaries

2.1 Notations

Throughout the paper, $\mathbb{N}$ is the set of non-negative integers and $\mathcal{H}$ is a real Hilbert space with scalar product $\langle\cdot,\cdot\rangle$, norm $\|\cdot\|$. Id denotes the identity operator on $\mathcal{H}$. The subdifferential of a proper function $h : \mathcal{H} \to]-\infty,+\infty]$ is the set–valued operator,

$$\partial h : \mathcal{H} \to 2^{\mathcal{H}},\ x \mapsto \big\{g \in \mathcal{H} | (\forall y \in \mathcal{H}),\ \langle y-x, g\rangle + h(x) \le h(y)\big\}.$$

$\Gamma_0(\mathcal{H})$ denotes the class of proper, lower semicontinuous, convex function from $\mathcal{H}$ to $]-\infty,+\infty]$. If $h \in \Gamma_0(\mathcal{H})$, then prox_h denotes the Moreau proximity operator [51], moreover, the Moreau envelope of index $\delta \in]0,+\infty[$ of h is the function,

$$^{\delta}h : x \mapsto \min_{y\in\mathcal{H}} h(y) + \frac{1}{2\delta}\|y-x\|^2,$$

and its gradient is δ^{-1}–Lipschitz continuous [50].

Let $A : \mathcal{H} \to 2^{\mathcal{H}}$ be a set–valued operator, then, the domain of A is $\mathrm{dom}A = \{x \in \mathcal{H} | Ax \neq \emptyset\}$; the range of A is $\mathrm{ran}A = \{y \in \mathcal{H} | \exists x \in \mathcal{H} : y \in Ax\}$; the graph of A is the set $\mathrm{gra}A = \{(x,y) \in \mathcal{H}^2 | y \in Ax\}$; the inverse of A is the operator whose graph is $\mathrm{gra}A^{-1} = \{(y,x) \in \mathcal{H}^2 | x \in A^{-1}y\}$; and its zeros set is $\mathrm{zer}A = \{x \in \mathcal{H} | 0 \in Ax\} = A^{-1}(0)$.

The resolvent of A is the operator $J_A = (\mathrm{Id}+A)^{-1}$, and the reflection operator associated to J_A is $R_A = 2J_A - \mathrm{Id}$.

Finally, denote ℓ^1_+ the set of summable sequences in $[0,+\infty[$, define the index set $[\![1,n]\!] = \{1,2,\cdots,n\}$ and through out the paper let $i \in [\![1,n]\!]$.

2.2 Non-expansive operators

Definition 2.1 (Non-expansive operator). An operator $T : \mathcal{H} \to \mathcal{H}$ is non-expansive if

$$\forall x, y \in \mathcal{H}, \ \|Tx - Ty\| \leq \|x - y\|.$$

For any $\alpha \in]0, 1[$, T is α–averaged if there exists a non-expansive operator R such that $T = \alpha R + (1 - \alpha)\mathrm{Id}$.

We denote $\mathcal{A}(\alpha)$ the class of α–averaged operators on $\mathcal{H}$, in particualr $\mathcal{A}(\frac{1}{2})$ is the class of firmly non-expansive operators, whose property is presented in Lemma 2.3.

Lemma 2.2. *Let $T : \mathcal{H} \to \mathcal{H}$ be an α–averaged operator, then $\frac{1}{2\alpha}(\mathrm{Id} - T) \in \mathcal{A}(\frac{1}{2})$.*

Proof. By definition, there exists a 1–Lipschitz continuous operator R such that $T = \alpha R + (1 - \alpha)\mathrm{Id}$, hence

$$\frac{1}{2\alpha}(\mathrm{Id} - T) = \frac{1}{2\alpha}\big(\mathrm{Id} - (\alpha R + (1 - \alpha)\mathrm{Id})\big) = \frac{1}{2}\big(\mathrm{Id} + (-R)\big) \in \mathcal{A}(\tfrac{1}{2}).$$

□

The following lemma gives some useful characterizations of firmly non-expansive operators.

Lemma 2.3. *The following statements are equivalent:*

(i) *T is firmly non-expansive;*

(ii) *$2T - \mathrm{Id}$ is non-expansive;*

(iii) *$\forall x, y \in \mathcal{H}, \ \|Tx - Ty\|^2 \leq \langle Tx - Ty, x - y\rangle$;*

(iv) *T is the resolvent of a maximal monotone operator A, i.e. $T = J_A$.*

Proof. For (i)-(iii), see [2, Chapter 4, Proposition 4.2]; For (i) ⇔ (iv), see [48]. □

The next lemma shows that α–averaged operators are closed under relaxations, convex combinations and compositions.

Lemma 2.4. *Let $(T_i)_{i\in[\![1,n]\!]}$ be a finite family of non-expansive operators from $\mathcal{H}$ to $\mathcal{H}$, $(\omega_i)_i \in]0, 1]^n$ and $\sum_i \omega_i = 1$, and let $(\alpha_i)_i \in]0, 1]^n$ such that, for every $i \in [\![1, n]\!]$, $T_i \in \mathcal{A}(\alpha_i)$. Then,*

(i) *$\big(\forall i \in [\![1, n]\!]\big) \ \big(\forall \lambda_i \in]0, \frac{1}{\alpha_i}[\big)$, $\mathrm{Id} + \lambda_i(T_i - \mathrm{Id}) \in \mathcal{A}(\lambda_i\alpha_i)$;*

(ii) *$T_1 \cdots T_n \in \mathcal{A}(\alpha)$, with*

$$\alpha = \frac{n}{n - 1 + \frac{1}{\max_{i\in[\![1,n]\!]} \alpha_i}};$$

(iii) *If $\bigcap_{i=1}^n \mathrm{fix}T_i \neq \emptyset$, then $\bigcap_{i=1}^n \mathrm{fix}T_i = \mathrm{fix}(T_1 \cdots T_n)$.*

Proof. (i) See [2, Chapter 4, Proposition 4.28]; (ii) See [2, Chapter 4, Proposition 4.32]; (iii) See [2, Chapter 4, Corollary 4.36]. □

Remark 2.5. For the composite operator $T_1 \cdots T_n$, a sharper bound of α can be obtained if $n = 2$,

$$\alpha = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2} \in]0, 1[,$$

see [53, Theorem 3].

Lemma 2.6. *Let $T : \mathcal{H} \to \mathcal{H}$ be non-expansive, then the set of fixed points* $\text{fix}T = \{x \in \mathcal{H} | x = Tx\}$ *is closed and convex.*

Proof. See [2, Chapter 4, Corollary 4.15]. □

Definition 2.7 (Quasi–Fejér sequences). Let Ω be a non-empty closed and convex subset of $\mathcal{H}$. A sequence $(x^k)_{k\in\mathbb{N}}$ is *quasi–Fejér monotone* with respect to Ω, *i.e.* there exists a summable sequence $(\varepsilon^k)_{k\in\mathbb{N}} \in \ell^1_+$ such that

$$\forall x \in \Omega,\ k \in \mathbb{N}, \quad \|x^{k+1} - x\| \leq \|x^k - x\| + \varepsilon^k.$$

2.3 Monotone operators

Definition 2.8 (Monotone operator). An operator A is monotone if

$$\big(\forall (x, u) \in \text{gra}A\big)\big(\forall (y, v) \in \text{gra}A\big),\ \langle x - y, u - v\rangle \geq 0,$$

it's moreover maximal monotone if gra$A$ is not strictly contained in the graph of any other monotone operator.

Definition 2.9 (Cocoercive operator). Let $\beta \in]0, +\infty[$, $B : \mathcal{H} \to \mathcal{H}$, then B is β–cocoercive if βT is firmly non-expansive, *i.e.*,

$$(\forall x, y \in \mathcal{H}),\ \beta\|Bx - By\|^2 \leq \langle Bx - By, x - y\rangle,$$

this indicates that B is $\frac{1}{\beta}$–Lipschitz continuous.

Lemma 2.10. *Let $A_1, A_2 : \mathcal{H} \to 2^{\mathcal{H}}$ be two maximal monotone operators, $\gamma \in]0, +\infty[$, and let $T = R_{\gamma A_1} R_{\gamma A_2}$, then*

(i) *T is non-expansive;*

(ii) $\frac{1}{2}(T + \text{Id}) = J_{\gamma A_1}(2J_{\gamma A_2} - \text{Id}) - J_{\gamma A_2} + \text{Id} \in \mathcal{A}(\frac{1}{2})$*;*

(iii) $(A_1 + A_2)^{-1}(0) = J_{\gamma A_2}(\text{fix}T)$*.*

Proof. See [19, Lemma 2.6]; □

Some properties of the subdifferentials are shown in the next lemma.

Lemma 2.11. *Let $f : \mathcal{H} \to]-\infty, +\infty[$ be a convex differentiable function, with $\frac{1}{\beta}$–Lipschitz continuous gradient, $\beta \in]0, +\infty[$, and let $h \in \Gamma_0(\mathcal{H})$. Then*

(i) $\beta\nabla f \in \mathcal{A}(\frac{1}{2})$*, i.e. is firmly non-expansive;*

(ii) $\text{Id} - \gamma\nabla f \in \mathcal{A}(\frac{\gamma}{2\beta})$ *for* $\gamma \in]0, 2\beta[$*;*

(iii) *∂h is maximal monotone;*

(iv) *The resolvent of ∂h is the proximity operator of h, i.e.* $\text{prox}_h = J_{\partial h}$*.*

Proof. (i) See [1, Baillon–Haddad theorem]; (ii) See [2, Chapter 4, Proposition 4.33]; (iii) See [60]; (iv) See [51]. □

2.4 Product Space

Let $(\omega_i)_i \in]0, +\infty]^n$, consider $\boldsymbol{\mathcal{H}} = \mathcal{H}^n$ endowed with the scalar product and norm defined as

$$\forall \boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{\mathcal{H}},\ \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \sum_{i=1}^{n} \omega_i \langle x_i, y_i \rangle,\ \|\boldsymbol{x}\| = \sqrt{\textstyle\sum_{i=1}^{n} \omega_i \|x_i\|^2}.$$

Define the non-empty closed convex set $\boldsymbol{\mathcal{S}} \subset \boldsymbol{\mathcal{H}}$, and its orthogonal complement $\boldsymbol{\mathcal{S}}^\perp \subset \boldsymbol{\mathcal{H}}$ which is a closed linear subspace. Let $\mathbf{Id}$ denotes the identity operator on $\boldsymbol{\mathcal{H}}$, then for $\forall \boldsymbol{z} = (z_i)_i \in \boldsymbol{\mathcal{H}}$, the project of $\boldsymbol{z}$ to $\boldsymbol{\mathcal{S}}$ is defined as

$$\boldsymbol{z}_{\boldsymbol{\mathcal{S}}} = P_{\boldsymbol{\mathcal{S}}}(\boldsymbol{z}).$$

Define $\iota_{\boldsymbol{\mathcal{S}}} : \boldsymbol{\mathcal{H}} \to]-\infty, +\infty]$ and $N_{\boldsymbol{\mathcal{S}}} : \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}$ the indicator function and the normal cone of the subspace $\boldsymbol{\mathcal{S}}$,

$$\iota_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \begin{cases} 0, & \text{if } \boldsymbol{x} \in \boldsymbol{\mathcal{S}}, \\ +\infty, & \text{otherwise}, \end{cases} \qquad N_{\boldsymbol{\mathcal{S}}}(\boldsymbol{x}) = \begin{cases} \boldsymbol{\mathcal{S}}^\perp, & \text{if } \boldsymbol{x} \in \boldsymbol{\mathcal{S}}, \\ \emptyset, & \text{otherwise}. \end{cases}$$

Since $\boldsymbol{\mathcal{S}}$ is non-empty closed and convex, it is straightforward that $N_{\boldsymbol{\mathcal{S}}}$ is maximal monotone.

For arbitary maximal monotone operators $A_i,\ i \in [\![1, n]\!]$ and B from $\mathcal{H} \to 2^{\mathcal{H}}$, let $\boldsymbol{\gamma} = (\gamma_i)_i \in]0, +\infty[^n$, and define $\boldsymbol{\gamma A} : \boldsymbol{\mathcal{H}} \to 2^{\boldsymbol{\mathcal{H}}}, \boldsymbol{x} = (x_i)_i \mapsto \bigtimes_{i=1}^{n} \gamma_i A_i x_i$, *i.e.* its graph is

$$\text{gra}\boldsymbol{\gamma A} = \bigtimes_{i=1}^{n} \text{gra}\gamma_i A_i = \Big\{ (\boldsymbol{x}, \boldsymbol{y}) \in \boldsymbol{\mathcal{H}}^2 | \boldsymbol{x} = (x_i)_i,\ \boldsymbol{y} = (y_i)_i,\ \forall i,\ y_i \in \gamma_i A_i x_i \Big\},$$

and $\boldsymbol{B} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}},\ \boldsymbol{x} = (x_i)_i \mapsto (Bx_i)_i$. Obviously, both $\boldsymbol{\gamma A}$ and $\boldsymbol{B}$ are maximal monotone on $\boldsymbol{\mathcal{H}}$ since $A_1, \cdots, A_n$ and B are maximal monotone.

3 Iteration–complexity bounds

In this section, we present the global iteration–complexity bounds of the inexact Krasnosel'skiĭ–Mann iteration [38, 45].

Definition 3.1 (Inexact Krasnosel'skiĭ–Mann iteration). Let $T : \mathcal{H} \to \mathcal{H}$ be a non-expansive operator such that $\text{fix}T \neq \emptyset$, let $\lambda_k \in]0, 1]$, then the inexact Krasnosel'skiĭ–Mann iteration of T is defined by:

$$\begin{aligned} z^{k+1} &= z^k + \lambda_k \big(T z^k + \varepsilon^k - z^k \big) \\ &= \big(\lambda_k T + (1 - \lambda_k)\text{Id} \big) z^k + \lambda_k \varepsilon^k \\ &= T_k z^k + \lambda_k \varepsilon^k, \end{aligned} \tag{3.1}$$

where ε^k is the error of approximating Tz^k. Define $T' = \text{Id} - T$ and the error of the iteration

$$e^k = (\text{Id} - T) z^k = T' z^k = \frac{z^k - z^{k+1}}{\lambda_k} + \varepsilon^k. \tag{3.2}$$

We start by collecting some useful results characterizing the above iteration.

Proposition 3.2. *Tollowing statements hold,*

(i) $T_k \in \mathcal{A}(\lambda_k)$ *if* $\lambda_k \in]0, 1[$*. If moreover* $T \in \mathcal{A}(\alpha)$*, then* $T_k \in \mathcal{A}(\lambda_k \alpha)$*;*

(ii) *For any* $z^\star \in \text{fix}T$*,*

$$z^\star \in \text{fix}T \iff z^\star \in \text{fix}T_k \iff z^\star \in \text{zer}T';$$

(iii) *If* $\big(\lambda_k(1-\lambda_k)\big)_{k\in\mathbb{N}} \notin \ell^1_+$ *and* $\big(\lambda_k\|\varepsilon^k\|\big)_{k\in\mathbb{N}} \in \ell^1_+$*, then,*

(1) e^k *converges strongly to* 0*.*

(2) *Sequence* $(z^k)_{k\in\mathbb{N}}$ *is quasi–Fejér monotone with respect to* $\mathrm{fix}T$*, and converges weakly to a point* $z^\star \in \mathrm{fix}T$*.*

Proof. (i) See Definition 2.1 and Lemma 2.4 (i); (ii) Straightforward; (iii) See [19, Lemma 5.1]. □

We are now in position to present the global iteration–complexity bounds for (3.1). The proofs are deferred to Appendix B. Define $\tau_k = \lambda_k(1-\lambda_k)$, $\underline{\tau} = \inf_{k\in\mathbb{N}} \tau_k$, $\overline{\tau} = \sup_{k\in\mathbb{N}} \tau_k$, and $\nu_1 = 2\sup_{k\in\mathbb{N}} \|T_k z^k - z^\star\| + \sup_{k\in\mathbb{N}} \lambda_k\|\varepsilon^k\|$, $\nu_2 = 2\sup_{k\in\mathbb{N}} \|e^k - e^{k+1}\|$.

Theorem 3.3 (Pointwise iteration–complexity bound). *Let* d_0 *be the distance from the starting point* z^0 *to the set of fixed points* $\mathrm{fix}T$*, if there holds,*

$$0 < \inf_{k\in\mathbb{N}} \lambda_k \le \sup_{k\in\mathbb{N}} \lambda_k < 1 \ \text{ and } \ \big((k+1)\|\varepsilon^k\|\big)_{k\in\mathbb{N}} \in \ell^1_+ \ , \tag{3.3}$$

define $C_1 = \nu_1 \sum_{j\in\mathbb{N}} \lambda_j\|\varepsilon^j\| + \nu_2\overline{\tau}\sum_{\ell\in\mathbb{N}}(\ell+1)\|\varepsilon^\ell\| < +\infty$*, then we have*

$$\|e^k\| \le \sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}}. \tag{3.4}$$

Moreover, if $\frac{1}{2} \le \lambda_k \le \sup_{k\in\mathbb{N}} \lambda_k < 1$ *is non-decreasing, then*

$$\|e^k\| \le \sqrt{\frac{d_0^2 + C_1}{\tau_k(k+1)}}, \tag{3.5}$$

Remark 3.4.

(i) Finding $z^\star \in \mathrm{fix}T$ is equivalent to finding a zero of T', see Proposition 3.2. Thus, Theorem 3.3 tells us that $O(1/\epsilon)$ iterations are needed for (3.1) to reach an ϵ–accurate in terms of the error criterion $\|T'z^k\|^2 \le \epsilon$.

(ii) For the case of first–order methods for solving smooth optimization problems, *i.e.* the gradient descent where T' is just the gradient, the obtained pointwise bound is the best–known complexity bound [52].

(iii) When the fixed point iteration (3.1) is exact, the sequence $(\|e^k\|)_{k\in\mathbb{N}}$ is non-increasing (Lemma A.5), hence we get

$$\|e^k\| \le \frac{d_0}{\sqrt{\sum_{j=0}^k \tau_j}},$$

which recovers the result of [22, Propositon 11].

When T is α–averaged operator, the above result still holds under a slight modification.

Corollary 3.5. *If* T *is* α*–averaged, then condition* (3.3) *changes to*

$$0 < \inf_{k\in\mathbb{N}} \lambda_k \le \sup_{k\in\mathbb{N}} \lambda_k < \frac{1}{\alpha} \ \text{ and } \ \big((k+1)\|\varepsilon^k\|\big)_{k\in\mathbb{N}} \in \ell^1_+ \ ,$$

and the conlusions of Theorem 3.3 still hold with $\tau_k = \lambda_k\big(\frac{1}{\alpha} - \lambda_k\big)$ *and* C_1 *defined accordingly.*

We now turn to the ergodic iteration–complexity bound of (3.1). For this, let's define $\Lambda_k = \sum_{j=0}^k \lambda_j$, and $\bar{e}^k = \frac{1}{\Lambda_k}\sum_{j=0}^k \lambda_j e^j$.

Theorem 3.6 (Ergodic iteration–complexity bound). *If* $C_2 = \sum_{k\in\mathbb{N}} \lambda_k \|\varepsilon^k\| < +\infty$*, then,*

$$\|\bar{e}^k\| \leq \frac{2(d_0 + C_2)}{\Lambda_k}.$$

In particular, if $\inf_{k\in\mathbb{N}} \lambda_k > 0$*, one obtains* $\|\bar{e}^k\| = O\big(1/(k+1)\big)$.

Again, this result holds when T is α–averaged, where now λ_k is allowed to vary in $]0, 1/\alpha]$.

From Theorem 3.3 and 3.6, it is immediate to get the convergence rate bounds on the sequence $\big(\|z^k - z^{k+1}\|\big)_{k\in\mathbb{N}}$ in the exact case. To lighten notation, let $v^k = z^k - z^{k+1}$ and $\bar{v}^k = \frac{1}{k+1}\sum_{j=0}^k v^j$.

Corollary 3.7. *Assume that* $\varepsilon^k = 0$ *for all* $k \in \mathbb{N}$.

(i) *If* $0 < \inf_{k\in\mathbb{N}} \lambda_k \leq \sup_{k\in\mathbb{N}} \lambda_k < 1$*, then*

$$\|v_k\| \leq \frac{d_0}{\sqrt{\underline{\tau}(k+1)}}.$$

(ii) *If* $\underline{\lambda} = \inf_{k\in\mathbb{N}} \lambda_k > 0$*, then*

$$\|\bar{v}^k\| \leq \frac{2d_0}{k+1}.$$

4 Local convergence rate

For many splitting algorithms applied to a range of optimization problems we consider in Section 7, see also the instructive example of gradient descent discussed at the end of this section, a typical convergence profile shows a global sublinear rate, and after a sufficiently large number of iterations, the algorithm enters a new regime where a local linear convergence takes over. This has also been observed by several authors, for instance with DRS or FBS to solve ℓ_1–minimization problems, see *e.g.* [24].

In this section, we study the rationale underlying this local linear convergence behavior. Our analysis relies on metric subregularity of the operator $T' = \mathrm{Id} - T$, where we recall that T is the fixed point operator, see Definition 3.1.

Definition 4.1 (Metric subregularity [26]). A set–valued mapping $F : \mathcal{H} \to 2^{\mathcal{H}}$ is called metrically subregular at $\tilde{z}$ for $\tilde{u} \in F(\tilde{z})$ if there exists $\kappa \geq 0$ along with neighbourhood $\mathcal{Z}$ of $\tilde{z}$ such that

$$d(z, F^{-1}\tilde{u}) \leq \kappa\, d(\tilde{u}, Fz), \quad \forall z \in \mathcal{Z}. \tag{4.1}$$

The infimum of κ for which this holds is the modulus of metric subregularity, denoted by $\mathrm{subreg}(F; \tilde{z}|\tilde{u})$. The absence of metric regularity is signaled by $\mathrm{subreg}(F; \tilde{z}|\tilde{u}) = +\infty$.

Metric subregularity implies that, for any $z \in \mathcal{Z}$, $d(\tilde{u}, Fz)$ is bounded below. The metric (sub)regularity of multifunctions plays a crucial role in modern variational analysis and optimization. These properties are a key to study the stability of solutions of generalized equations, see the dedicated monograph [26].

Let's specialize this definition to the operator T' and $\tilde{u} = 0$. T' is single–valued and $T'^{-1}(0) = \mathrm{fix}T$. Thus if T' is metrically subregular at some $z^\star \in \mathrm{fix}T$ for 0, then from (4.1) we have

$$d(z, \mathrm{fix}T) \leq \kappa \|T'z\|, \quad \forall z \in \mathcal{Z}.$$

Metric subregularity implies that (4.1) gives an estimate for how far a point z is from being the fixed point set of T in terms of the residual $\|z - Tz\|$. This is the rationale behind using such a regularity assumption on the operator T' to quantify the convergence rate on $d(z^k, \mathrm{fix}T)$. Thus, starting from $z^0 \in \mathcal{H}$, and by virtue of Theorem 3.3, one can recover a $O(1/\sqrt{k})$ rate on $d(z^k, \mathrm{fix}T)$. In fact, we can do even better as shown in the following theorem. We use the shorthand notation $d_k = d(z^k, \mathrm{fix}T)$.

Theorem 4.2 (Local convergence rate). *Let $z^\star \in \mathrm{fix}T$, assume that T' is metrically subregular at $z^\star$ with neighbourhood $\mathcal{Z}$ of $z^\star$, let $\kappa > \mathrm{subreg}(T'; z^\star|0)$, $\lambda_k \in]0,1]$. Assume that $C_2 = \sum_{k\in\mathbb{N}} \lambda_k \|\varepsilon^k\|$ is sufficiently small, and there exists a ball $\mathbb{B}_a(z^\star)$, $a \geq 0$, such that*

$$\mathbb{B}_{(a+C_2)}(z^\star) \subseteq \mathcal{Z}.$$

Then for any starting point $z^0 \in \mathbb{B}_a(z^\star)$, we have for all $k \in \mathbb{N}$,

$$d_{k+1}^2 \leq \zeta_k \, d_k^2 + c_k,$$

where $\zeta_k = \begin{cases} 1 - \frac{\tau_k}{\kappa^2}, & \text{if } \tau_k/\kappa^2 \in]0,1] \\ \frac{\kappa^2}{\kappa^2+\tau_k}, & \text{otherwise} \end{cases} \in [0,1[$, $c_k = \nu_1 \lambda_k \|\varepsilon^k\|$. *Moreover,*

(i) *d_k converges. If $(\tau_k)_{k\in\mathbb{N}} \notin \ell^1_+$, it converges to 0;*

(ii) *Let $\chi_k = \prod_{j=0}^k \zeta_j$, if $\chi = \limsup_{k\to+\infty} \sqrt[k]{\chi_k} < 1$, then*

  (1) $\left(d_k^2\right)_{k\in\mathbb{N}} \in \ell^1_+$;

  (2) *If $\varepsilon^k = 0$, $\lim_{k\to+\infty} \sqrt[k]{d_k} < 1$, which is R–linear convergence* [2].

(iii) *If $0 < \inf_{k\in\mathbb{N}} \lambda_k \leq \sup_{k\in\mathbb{N}} \lambda_k < 1$, then there exists $\zeta \in (0,1)$ such that*

$$d_{k+1}^2 \leq \zeta^k \left(d_0^2 + \textstyle\sum_{j=0}^k \zeta^{-j+1} c_j \right).$$

(iv) *If the set of fixed points $\mathrm{fix}T = \{z^\star\}$ is a singleton, then*

  (1) *$\|z^k - z^\star\|$ converges. If $(\tau_k)_{k\in\mathbb{N}} \notin \ell^1_+$, then $\|z^k - z^\star\|$ converges to 0;*

  (2) *If $\limsup_{k\to+\infty} \sqrt[k]{\chi_k} < 1$, then $\left(\|z^k - z^\star\|^2\right)_{k\in\mathbb{N}} \in \ell^1_+$, and $\lim_{k\to+\infty} \sqrt[k]{\|z^k - z^\star\|} < 1$ if $\varepsilon^k = 0$;*

  (3) *If $0 < \inf_{k\in\mathbb{N}} \lambda_k \leq \sup_{k\in\mathbb{N}} \lambda_k < 1$, then there exists $\zeta \in (0,1)$ such that*

$$\|z^{k+1} - z^\star\|^2 \leq \zeta^k \left(\|z^0 - z^\star\|^2 + \textstyle\sum_{j=0}^k \zeta^{-j+1} c_j \right),$$

  and linear convergence rate is obtained if $c_k = o(\zeta^k)$.

[2] A sequence $(x^k)_{k\in\mathbb{N}}$ is said to converge Q–linearly to $\tilde{x}$ if there exists a constant $r \in]0,1[$ such that $\|x^{k+1} - \tilde{x}\|/\|x^k - \tilde{x}\| \leq r$, and $(x^k)_{k\in\mathbb{N}}$ is said to converge R–linearly to $\tilde{x}$ if $\|x^k - \tilde{x}\| \leq \sigma_k$ and $(\sigma_k)_{k\in\mathbb{N}}$ converges Q–linearly to 0.

For the sake of clarity, the different regimes exhibited by Theorem 4.2 are summarized in the diagram of Figure 1.

Remark 4.3.

(i) For simplicity, suppose the iteration is exact, and let $z^\star \in \text{fix}T$ such that $d_k = \|z^k - z^\star\|$. Then we have

$$\|e^k\|^2 = \|z^k - z^\star + Tz^\star - Tz^k\|^2 \leq 4d_k^2 \leq 4\zeta^k d_0^2,$$

which means that locally, $\|e^k\|$ also converges linearly to 0.

(ii) As far as the claim in Theorem 4.2 (iii) is concerned, if there exists $\xi \in]0,1]$ such that $c_k = O(\xi^k)$, then

(1) If $\xi < \zeta$, then $d_{k+1}^2 = O(\zeta^k)$;

(2) If $\xi = \zeta$, then $d_{k+1}^2 = O(k\zeta^k) = o\big((\zeta+\delta)^k\big)$, $\delta > 0$;

(3) If $\xi > \zeta$, then $d_{k+1}^2 = o(\xi^k)$.

Corollary 4.4. *If T is α–averaged, then Theorem 4.2 holds substituting $\lambda_k\alpha$ for λ_k.*

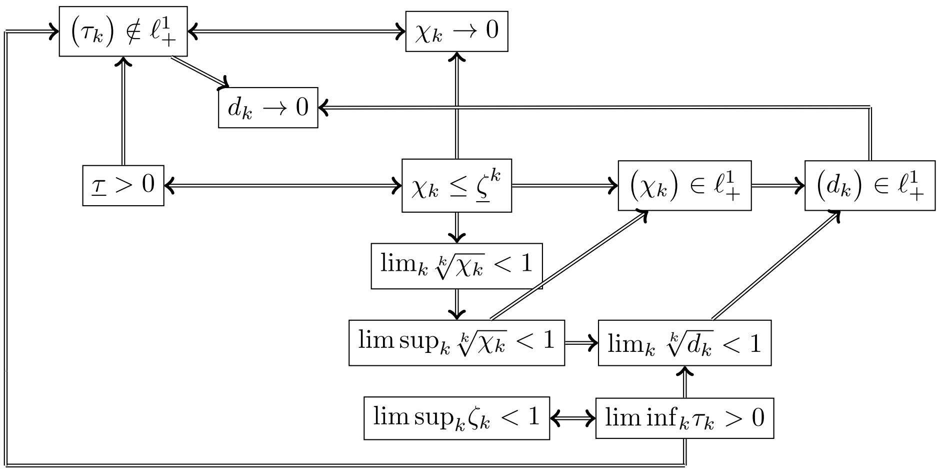


Figure 1: Diagram of the result of Theorem 4.2.

Remark 4.5. Equivalent characterizations of metric subregularity can be given, for instance in terms of derivative criteria. In particular, as T' is single–valued, metric subregularity of T' holds if T' is differentiable on a neighbourhood of $z^\star$ with non-singular derivatives at z around $z^\star$, and the operator norms of their inverses are uniformly bounded [25, Theorem 1.2]. Computing the metric regularity modulus κ is however far from obvious in general even for the differentiable case and these details are left to a future work.

Example: Gradient Descent

The aim of this example is to illustrate the meaning of the metric subregularity assumption on T' and the entailed linear rate of Theorem 4.2 when minimizing a convex smooth function whose gradient is Lipschitz continuous. Let's consider the minimization of the function

$$f : x \in \mathbb{R}^2 \mapsto f(x) = f_{\delta_1}(x_1) + f_{\delta_2}(x_2)\ ,$$

where f_δ is the Moreau enveloppe of $x \in \mathbb{R} \mapsto |x|$ of order $\delta > 0$

$$f_\delta : x \in \mathbb{R} \mapsto \begin{cases} \frac{x^2}{2\delta}, & \text{if } |x| \le \delta \\ |x| - \frac{\delta}{2}, & \text{if } |x| > \delta. \end{cases}$$

Denote $\delta_{\min} = \min(\delta_1, \delta_2)$ and $\delta_{\max} = \max(\delta_1, \delta_2)$. It is easy to see that f is a continuous differentiable convex function, whose gradient ∇f is $1/\delta_{\min}$–Lipschitz continuous. Moreover, it has a unique minimizer at 0, and is locally $1/\delta_{\max}$–strongly convex in $]-\delta_{\min}, \delta_{\min}[^2$.

For simplicity, consider the non-relaxed gradient descent method for minimizing f with constant step–size, *i.e.*

$$x^{k+1} = x^k - \gamma \nabla f(x^k),$$

where $\gamma \in]0, 2\delta_{\min}[$.

This can be cast in our above framework by setting $T = \mathrm{Id} - \gamma \nabla f$, where obviously, T is $\frac{\gamma}{2\delta_{\min}}$–averaged (Lemma 2.11). Hence $T' = \mathrm{Id} - T = \gamma \nabla f$ which is continuously differentiable on any neighborhood $\mathcal{Z} \subset]-\delta_{\min}, \delta_{\min}[^2$. For any $x \in \mathcal{Z}$, the Jacobian of T' is just $\gamma \nabla^2 f(x)$, where $\nabla^2 f(x)$ is the Hessian of f at x, which is non-singular and its inverse is uniformly bounded by $\delta_{\max}$ from local strong convexity. Thus, by virtue of [26, Theorem 4B.1], T' is metrically regular, hence subregular, and the metric regularity modulus κ is precisely $\delta_{\max}/\gamma$. In fact we could have anticipated this directly from the local strong monotonicity of ∇f. Specializing the rate of Theorem 4.2, we get

$$\zeta = 1 - \frac{t(2-t)}{\mathrm{cnd}^2} \in [0, 1[,$$

where we set $t = \gamma/\delta_{\min} \in]0, 2[$, and $\mathrm{cnd} = \delta_{\max}/\delta_{\min}$ can be seen as the condition number of the Hessian of f in $]-\delta_{\min}, \delta_{\min}[^2$. It is clear that the best rate[3] is attained for $t = 1$, *i.e.* $\gamma = \delta_{\min}$.

The observed and theoretical convergence profiles of $\|e^k\| = \|x^k - x^{k+1}\|$ are illustrated in Figure 2 where gradient descent was run with $\gamma \in \{\delta_{\min}/2, \delta_{\min}\}$. As predicted by our result, the convergence profile exhibits two regimes, a global sublinear one, and then a local linear one. This will be confirmed on more elaborated high–dimensional numerical experiments in Section 7.

5 Applications

In this section, we apply the obtained results to conduct quantitative convergence analysis of a class of monotone operator splitting methods proposed in the literature, and mainly focus on the global iteration–complexity bounds. As stated in the introduction, we will rely on the fact that all the considered iterative schemes (GFB/FBS, DRS/ADMM and PDS) can be cast as Krasnosel'skiĭ–Mann iteration. Furthermore, based on the structure of the corresponding monotone inclusion problem, we also derive specific error criteria which can serve termination tests. The proofs of the main theorems are deferred to Appendix D.

5.1 Generalized Forward–Backward splitting

The GFB algorithm is proposed in [57] to address the monotone inclusion problem (1.1) when $L_i = \mathrm{Id}, i \in [\![1, n]\!]$. Details of the algorithm are given in Algorithm 1. By lifting the problem into the product space, the method achieves the full splitting of the A_i's, and like the classic Forward–Backward splitting method [21], the cocoercive operator B is directly applied to the minimizer.

[3] One may observe that γ can be modified locally, once the iterates enter the appropriate neighbourhood, to get the ususal optimal linear rate of gradient descent with a strongly convex objective. But this is not our aim here.

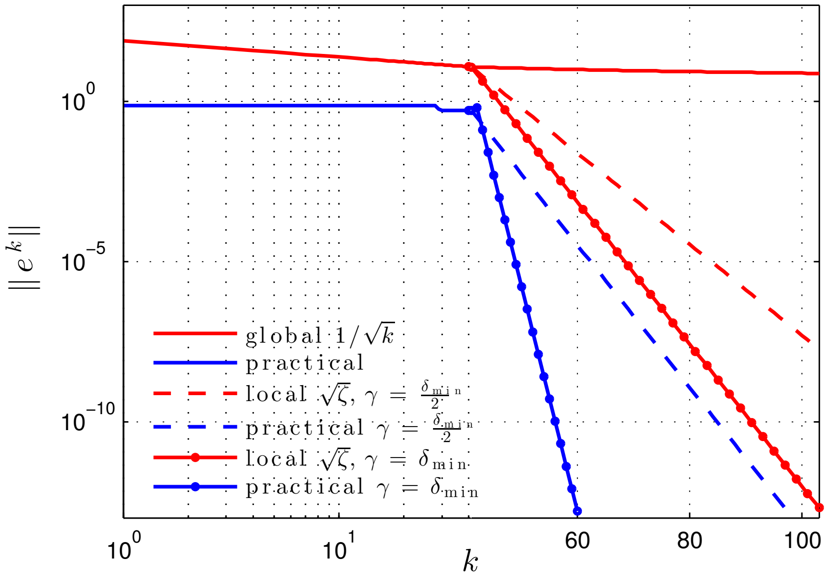


Figure 2: Global and local convergence profiles of gradient descent to minimize f with $(\delta_{\min}, \delta_{\max}) = (1, 1.25)$. For $\gamma = \delta_{\min}/2$, the observed and theoretical rates are respectively 0.60 and 0.72. For $\gamma = \delta_{\min}$, they are 0.20 and 0.60 respectively.

It can be observed that when $n = 1$, GFB recovers FBS, and when $B = 0$, GFB recovers the DRS in the product space. In the literature, the convergent property of the FBS and DRS has been extensively studied [5, 16, 21, 30, 65, 67, 33]. Hence in this section, we mainly focus on the GFB method.

In the following context, we first briefly recall the product space and the fixed point formulation of GFB ([57, Section 4]), then establish two types of iteration–complexity bounds for GFB based on the iteration itself and the structure of the monotone inclusion problem (1.1).

Algorithm 1: Generalized Forward–Backward splitting algorithm

---

Input: Let $(\omega_i)_i \in]0,1[^n$ such that $\sum_{i=1}^n \omega_i = 1$, $\gamma \in]0, 2\beta[$, $\lambda_k \in]0, \frac{4\beta-\gamma}{2\beta}[$. Let $(\varepsilon_1^k)_{k\in\mathbb{N}}$, $(\varepsilon_{2,i}^k)_{k\in\mathbb{N}}$, $i \in [\![1, n]\!]$ be absolutely summable sequences in $\mathcal{H}$.

Initial: $k = 0$, $\boldsymbol{z}^0 = (z_i^0)_i$, $x^0 = \sum_{i=1}^n \omega_i z_i^0$;

repeat

  for $i = 1$ **to** n **do**

    $z_i^{k+1} = z_i^k + \lambda_k\big(J_{\frac{\gamma}{\omega_i}A_i}(2x^k - z_i^k - \gamma Bx^k + \varepsilon_1^k) + \varepsilon_{2,i}^k - x^k\big)$;

  $x^{k+1} = \sum_{i=1}^n \omega_i z_i^{k+1}$;

  $k = k + 1$;

until *convergence*;

From Algorithm 1 we can obtain the product space of GFB. Let $\boldsymbol{\mathcal{S}} = \big\{\boldsymbol{x} = (x_i)_i \in \boldsymbol{\mathcal{H}} | x_1 = \cdots = x_n\big\}$ and its orthogonal complement $\boldsymbol{\mathcal{S}}^\perp \subset \boldsymbol{\mathcal{H}} = \big\{\boldsymbol{x} = (x_i)_i \in \boldsymbol{\mathcal{H}} | \sum_{i=1}^n \omega_i x_i = 0\big\}$. Now define the canonical isometry,

$$\boldsymbol{C} : \mathcal{H} \to \boldsymbol{\mathcal{S}},\ x \mapsto (x, \cdots, x),$$

then according to Subsection 2.4, we have

$$P_{\mathcal{S}}(\boldsymbol{z}) = \boldsymbol{C}\Big(\textstyle\sum_{i=1}^n \omega_i z_i\Big),$$

clearly $P_{\mathcal{S}}$ is self–adjoint, and the reflection operator is $R_{\mathcal{S}} = 2P_{\mathcal{S}} - \mathbf{Id}$. Define operator $\boldsymbol{B}_{\mathcal{S}} = \boldsymbol{B} \circ P_{\mathcal{S}}$, $J_{\boldsymbol{\gamma A}} = \big(J_{\gamma_i A_i}\big)_i$ and $R_{\boldsymbol{\gamma A}} = 2J_{\boldsymbol{\gamma A}} - \mathbf{Id}$.

5.1.1 Fixed point formulation

Define $\boldsymbol{T}_{1,\boldsymbol{\gamma}} = \frac{1}{2}\big(R_{\boldsymbol{\gamma A}} R_{\mathcal{S}} + \mathbf{Id}\big)$, $\boldsymbol{T}_{2,\gamma} = \mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}}$. Then we have the following proposition.

Proposition 5.1. *The following statements hold:*

(i) *The composed operator $\boldsymbol{T}_{1,\boldsymbol{\gamma}} \circ \boldsymbol{T}_{2,\gamma}$ is $\frac{2\beta}{4\beta-\gamma}$–averaged;*

(ii) *For any $\boldsymbol{\gamma} \in]0, +\infty[^n$, there exists a maximal monotone operator $\boldsymbol{A}'_{\boldsymbol{\gamma}} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ such that $\boldsymbol{T}_{1,\boldsymbol{\gamma}} = J_{\boldsymbol{A}'_{\boldsymbol{\gamma}}}$.*

Proof. (i) Lemma 2.10, Remark 2.5; (ii) See [57, Proposition 4.15]. □

If let $u_i^{k+1} = J_{\frac{\gamma}{\omega_i}A_i}(2x^k - z_i^k - \gamma B x^k)$, $i \in [\![1, n]\!]$, then the iterates of Algorithm 1 can be reformulated to the following:

$$\begin{cases} \text{For } i = 1, \cdots, n \\ \quad \left\lfloor \begin{array}{l} u_i^{k+1} = J_{\frac{\gamma}{\omega_i}A_i}\big(2x^k - z_i^k - \gamma B x^k\big), \\ z_i^{k+1} = z_i^k + \lambda_k\big(u_i^{k+1} + \varepsilon_i^k - x^k\big), \end{array}\right. \\ x^{k+1} = \sum_{i=1}^n \omega_i z_i^{k+1}, \end{cases}$$

where $\varepsilon_i^k = \big(J_{\frac{\gamma}{\omega_i}A_i}(2x^k - z_i^k - \gamma B x^k + \varepsilon_1^k) + \varepsilon_{2,i}^k\big) - u_i^{k+1}$. Let $\boldsymbol{\varepsilon}_1^k = \boldsymbol{C}(\varepsilon_1^k)$, $\boldsymbol{\varepsilon}_2^k = (\varepsilon_{2,i}^k)_i$, then we can obtain the fixed point iteration of GFB.

Proposition 5.2 (Fixed point formulation of GFB). *Let $\boldsymbol{\gamma} = \big(\frac{\gamma}{\omega_i}\big)_i$, then Algorithm 1 is equivalent to the following inexact relaxed fixed point iteration,*

$$\begin{cases} \boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k\big(\boldsymbol{T}_{1,\boldsymbol{\gamma}}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k - \boldsymbol{z}^k\big), \\ x^{k+1} = \sum_{i=1}^n \omega_i z_i^{k+1}. \end{cases} \tag{5.1}$$

To lighten the notation, let $\boldsymbol{T}_\gamma = \boldsymbol{T}_{1,\boldsymbol{\gamma}} \circ \boldsymbol{T}_{2,\gamma}$, then (3.2) can be written as

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k(\boldsymbol{T}_\gamma \boldsymbol{z}^k + \boldsymbol{\varepsilon}^k - \boldsymbol{z}^k) \tag{5.2}$$

where $\boldsymbol{\varepsilon}^k = \big(\boldsymbol{T}_{1,\gamma}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k\big) - \boldsymbol{T}_\gamma \boldsymbol{z}^k$. Obviously, this is exactly the form of (3.1). Therefore, the convergence of GFB is guaranteed by Theorem 3.3 (see also [57, Theorem 4.1]), and the iteration–complexity bounds are as same as in Theorem 3.3 and 3.6. Moreover, Proposition 5.1 indicates that we can establish certain iteration–complexity bounds for GFB method based on the composed structure of the fixed point operator $\boldsymbol{T}_\gamma$. Therefore, in the following context, we provide two forms of generalization of iteration–complexity bounds for the GFB method, one is based on the property of the fixed point operator $\boldsymbol{T}_\gamma$, the other is based on the GFB algorithm 1 itself. We assume in the following that sequence $(\lambda_k)_{k\in\mathbb{N}} \in [\frac{1}{2\alpha}, \frac{1}{\alpha}[$ is non-decreasing.

5.1.2 Iteration–complexity bounds

From Proposition 5.1 we have $J_{\boldsymbol{A}'_\gamma} \in \mathcal{A}(\frac{1}{2})$ and $\mathbf{Id} - \gamma \boldsymbol{B}_{\boldsymbol{\mathcal{S}}} \in \mathcal{A}(\frac{\gamma}{2\beta})$, and also the following equivalence holds

$$\boldsymbol{z}^\star \in \operatorname{fix}\boldsymbol{T}_\gamma \iff \boldsymbol{z}^\star \in \operatorname{zer}\Big(\frac{1}{\gamma}\boldsymbol{A}'_\gamma + \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\Big).$$

We can thus establish the first type of iteration–complexity bounds for GFB.

Define $\boldsymbol{v}^k = \boldsymbol{z}^k - \boldsymbol{z}^{k+1} + \lambda_k \boldsymbol{\varepsilon}^k$, $v^k = \sum_i \omega_i v_i$, $\boldsymbol{p}^{k+1} = \frac{1}{\lambda_k}\big(\boldsymbol{z}^{k+1} - (1-\lambda_k)\boldsymbol{z}^k\big) - \boldsymbol{\varepsilon}^k$, $\boldsymbol{g}^{k+1} = \frac{1}{\gamma}\boldsymbol{z}^k - \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\boldsymbol{z}^k - \frac{1}{\gamma}\boldsymbol{p}^{k+1}$.

Proposition 5.3 (Pointwise iteration–complexity bounds). *We have $\boldsymbol{v}^k = \lambda_k \boldsymbol{e}^k$, $\boldsymbol{g}^{k+1} \in \frac{1}{\gamma}\boldsymbol{A}'_\gamma \boldsymbol{p}^{k+1}$, under the conditions of Theorem 3.3, the following statements hold:*

(i)

$$\|\boldsymbol{v}^k\| \le \sqrt{\frac{\lambda_k(d_0^2 + C_1)}{(\frac{1}{\alpha} - \lambda_k)(k+1)}},\ \|v^k\| \le \sqrt{\frac{\lambda_k(d_0^2 + C_1)}{(\frac{1}{\alpha} - \lambda_k)(k+1)}},$$

(ii)

$$d\Big(0, \frac{1}{\gamma}\boldsymbol{A}'_\gamma \boldsymbol{p}^{k+1} + \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\boldsymbol{p}^{k+1}\Big) \le \frac{1}{\gamma}\sqrt{\frac{d_0^2 + C_1}{\tau_k(k+1)}}.$$

Let $\bar{\boldsymbol{v}}^k = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{v}^j, \bar{v}^k = \sum_{i=1}^n \omega_i \bar{v}_i$, $\bar{\boldsymbol{z}}^k = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{z}^j$, $\bar{\boldsymbol{p}}^{k+1} = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{p}^{j+1}$, $\bar{\boldsymbol{g}}^{k+1} = \frac{1}{\gamma}\bar{\boldsymbol{z}}^k - \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\bar{\boldsymbol{z}}^k - \frac{1}{\gamma}\bar{\boldsymbol{p}}^{k+1}$.

Proposition 5.4 (Ergodic iteration–complexity bounds). *Under the conditions of Theorem 3.6, the following statements hold:*

(i)

$$\|\bar{\boldsymbol{v}}^k\| \le \frac{2(d_0 + C_2)}{k+1},\ \|\bar{v}^k\| \le \frac{2(d_0 + C_2)}{k+1},$$

(ii)

$$\|\bar{\boldsymbol{g}}^{k+1} + \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\bar{\boldsymbol{p}}^{k+1}\| \le \frac{2(d_0 + C_2)}{\gamma\lambda_0(k+1)}.$$

where C_2 is the same as in Theorem 3.6.

Similarly to Corollary 3.7, pointwise convergence rates of $O\big(1/\sqrt{k}\big)$ can be obtained for the sequences $\|\boldsymbol{z}^k - \boldsymbol{z}^{k+1}\|$ and $\|x^k - x^{k+1}\|$, and $O(1/k)$ for their ergodic counterparts.

5.1.3 Iteration–complexity bounds of the monotone inclusion (1.1)

We now develop the iteration–complexity bounds of GFB for the monotone inclusion (1.1). This can serve as a termination criterion of GFB in practice.

Let $e_i^k = x^k - u_i^{k+1} = \frac{1}{\lambda_k}(z_i^k - z_i^{k+1}) + \varepsilon_i^k$, $i \in [\![1,n]\!]$, $g^{k+1} = \frac{1}{\gamma}x^k - Bx^k - \frac{1}{\gamma}\sum_i \omega_i u_i^{k+1}$.

Proposition 5.5 (Pointwise iteration–complexity bound). *We have $g^{k+1} \in \sum_i A_i u_i^{k+1}$, moreover, under the assumptions of Theorem 3.3,*

$$d\Big(0, \textstyle\sum_i A_i u_i^{k+1} + B\big(\sum_i \omega_i u_i^{k+1}\big)\Big) \le \frac{1}{\gamma}\sqrt{\frac{d_0^2 + C_1}{\tau_k(k+1)}}.$$

Define $\bar{u}_i^{k+1} = \frac{1}{k+1}\sum_{j=0}^k u_i^{j+1},\ i \in [\![1,n]\!],\ \bar{x}^k = \frac{1}{k+1}\sum_{j=0}^k x^j$ and $\bar{g}^{k+1} = \frac{1}{\gamma}\bar{x}^k - B\bar{x}^k - \frac{1}{\gamma}\sum_i \omega_i\bar{u}_i^{k+1}$.

Proposition 5.6 (Ergodic iteration–complexity bound). *Under the assumptions of Theorem 3.6, we have*

$$\|\bar{g}^{k+1} + B(\textstyle\sum_i \omega_i \bar{u}_i^{k+1})\| \le \frac{2(d_0 + C_2)}{\gamma\lambda_0(k+1)}.$$

where C_2 is the same in Theorem 3.6.

Remark 5.7.

(i) Theorem 5.5 indicates that GFB provides an ϵ–accurate solution in at most $O(1/\epsilon)$ iterations for the termination criterion $d\big(0, \sum_i A_i u_i^k + B(\sum_i \omega_i u_i^k)\big)^2 \le \epsilon$. This can then be viewed as a generalization of the best–known complexity bounds of the gradient descent method [52].

(ii) The product space $\boldsymbol{\mathcal{S}}$ of GFB is defined by $\boldsymbol{\mathcal{S}} = \{\boldsymbol{x} \in \boldsymbol{\mathcal{H}} | x_1 = \cdots = x_n\} \subset \boldsymbol{\mathcal{H}}$. In [7], the result is generalised to any subspace $\boldsymbol{\mathcal{V}} \subset \boldsymbol{\mathcal{H}}$. In fact, if we redefine $\boldsymbol{T}_{1,\gamma} = \frac{1}{2}\big(R_{\gamma\boldsymbol{A}}R_{\boldsymbol{\mathcal{V}}} + \mathbf{Id}\big)$, $\boldsymbol{T}_{2,\gamma} = \mathbf{Id} - \gamma P_{\boldsymbol{\mathcal{V}}}\boldsymbol{B}P_{\boldsymbol{\mathcal{V}}}$, then (5.1) recovers the Forward–Douglas–Rachford splitting (FDRS) method proposed in [7, Section 4]. Moreover, $\boldsymbol{T}_{2,\gamma} \in \mathcal{A}(\frac{\gamma}{2\beta})$ still holds ([7, Proposition 4.1]), therefore, FDRS method obeys the iteration–complexity bounds established in Section 3;

(iii) Using the transportation formula for cocoercive operators in [64, Lemma 2.8], one can write the non-relaxed and exact GFB method into the HPE framework introduced there.

However, there are two important aspects we would like to point out. First, the HPE framework is ”conceptual” in the sense that though one can cast GFB into the HPE framework, the implementable iterative scheme cannot be recovered from HPE. Secondly, to establish the iteration–complexity bounds for GFB under the HPE framework, no relaxation nor errors are handled, and the iteration–complexity bound would be non-uniform [49][4].

5.2 Douglas–Rachford splitting and ADMM

As mentioned in the introduction, the Douglas–Rachford splitting method [44] can be applied to solve (1.1) when $B = 0,\ n = 2$ and $L_i = \mathrm{Id},\ i = 1, 2$.

Algorithm 2: Douglas–Rachford splitting method

Input: $\gamma > 0$, $\lambda_k \in]0, 2]$, let $(\varepsilon_1^k)_{k\in\mathbb{N}}$, $(\varepsilon_2^k)_{k\in\mathbb{N}}$ be two absolutely summable sequences in $\mathcal{H}$.
Initial: $k = 0$, $z^0 \in \mathcal{H}$, $x^0 = J_{\gamma A_2}(z^0)$;
repeat
  $z^{k+1} = z^k + \lambda_k\big(J_{\gamma A_1}(2x^k - z^k) + \varepsilon_1^k - x^k\big)$;
  $x^{k+1} = J_{\gamma A_2}(z^{k+1}) + \varepsilon_2^{k+1}$;
  $k = k + 1$;
until *convergence*;

If for instance $L_1 : \mathcal{H} \to \mathcal{G}$ is some bounded linear operator, then the problem can be solved by the Alternating Direction method of Multipliers (ADMM) [30], which is applying DRS method to the dual formulation of the problem [28].

[4]Let $(x^k, v^k) \in \mathrm{gra}A$ be a sequence generated by an iterative method for solving the monotone inclusion problem $0 \in Ax$, then the non-uniform iteration–complexity bound means that for every $k \in \mathbb{N}$, there exists a $j \le k$ such that $\|v^j\| = O(1/\sqrt{k})$.

Define

$$T=\frac{1}{2}\big(R_{\gamma A_1}R_{\gamma A_2}+\mathrm{Id}\big),\quad \varepsilon^k=\Big(\frac{1}{2}\big(R_{\gamma A_1}(R_{\gamma A_2}\cdot+2\varepsilon_2^k)+\mathrm{Id}\big)z^k-Tz^k\Big)+\varepsilon_1^k,$$

then the fixed point formulation of the DRS with respect to z^k is exactly of the form (3.1). Moreover, $T\in\mathcal{A}(\frac{1}{2})$ is firmly non-expansive and $(\varepsilon_k)_{k\in\mathbb{N}}\in\ell_+^1$ owing to the summability of $(\varepsilon_1^k)_{k\in\mathbb{N}}$, $(\varepsilon_2^k)_{k\in\mathbb{N}}$. Therefore, the DRS method obeys the iteration–complexity bounds established in Section 3.

Next, we turn to the corresponding monotone inclusion problem (1.1), and develop a corresponding criterion similiar to Theorem 5.5, 5.6. Define $u^{k+1}=J_{\gamma A_1}(2x^k-z^k)$, $v^{k+1}=J_{\gamma A_2}(z^{k+1})$, and let $g^{k+1}=\frac{1}{\gamma}(2x^k-z^k-u^{k+1}+z^{k+1}-v^{k+1})$.

Proposition 5.8. *We have $g^{k+1}\in A_1u^{k+1}+A_2v^{k+1}$, and if $0<\inf_{k\in\mathbb{N}}\lambda_k\le\sup_{k\in\mathbb{N}}\lambda_k<2$, $(\varepsilon_1^k)_{k\in\mathbb{N}},(\varepsilon_2^k)_{k\in\mathbb{N}}\in\ell_+^1$,*

$$d\Big(0,A_1u^{k+1}+A_2v^{k+1}\Big)\le\frac{1+\lambda_k}{\gamma}\sqrt{\frac{d_0^2+C_1}{\underline{\tau}(k+1)}}+c_k, \tag{5.3}$$

where $c_k=\frac{1}{\gamma}\big((2+\lambda_k)\|\varepsilon_2^k\|+\|\varepsilon_1^k\|\big)$, d_0 and C_1 are as similar as defined in Theorem 3.3.

Remark 5.9.

(i) The summability assumption of $(\varepsilon_1^k)_{k\in\mathbb{N}}$ and $(\varepsilon_1^k)_{k\in\mathbb{N}}$ implies $(c_k)_{k\in\mathbb{N}}$ is summable too, hence decays faster than $1/k$, which means the right hand side of (5.3) is dominated by the first term.

However, to ensure the convergence of the DRS method, one only needs $\big(\lambda_k\|\varepsilon^k\|\big)_{k\in\mathbb{N}}$ to be summable, which does not necesary mean that $(\varepsilon_1^k)_{k\in\mathbb{N}}$ and $(\varepsilon_1^k)_{k\in\mathbb{N}}$ should be summable. In [18, Remark 5.7], an example is provided where $\sum_{k\in\mathbb{N}}\|\varepsilon^k\|$ may diverge while $\big(\lambda_k\|\varepsilon^k\|\big)_{k\in\mathbb{N}}$ is summable. Suppose $\|\varepsilon^k\|\le\big(1+\sqrt{1-1/k}\big)/k^q$, $q\in]0,1]$, and $\lambda_k=\big(1-\sqrt{1-1/k}\big)/2$, then it can be verified that $\sum_{k\in\mathbb{N}}\|\varepsilon^k\|$ diverges but $\sum_{k\in\mathbb{N}}\lambda_k\|\varepsilon^k\|<+\infty$ and $\sum_{k\in\mathbb{N}}\lambda_k(2-\lambda_k)=+\infty$.

(ii) The obtained result can be easily extended to the ADMM method, where we can prove that the sequences generated by the ADMM iteration converge at the rate of $O\big(1/\sqrt{k}\big)$ pointwisely, a similar result under different metric is presented in [34].

5.3 Vũ's primal–dual splitting

In [68], a more general monotone problem is considered. Let $\mathcal{H}$ be a real Hilbert space, $C:\mathcal{H}\to 2^{\mathcal{H}}$ is maximal monotone, $B:\mathcal{H}\to\mathcal{H}$ is μ–cocoercive for some $\mu\in]0,+\infty[$. n is a strictly positive integer, let $(\omega_i)_i\in]0,1[^n$ such that $\sum_i\omega_i=1$. For every $i\in[\![1,n]\!]$, let $\mathcal{G}_i$ be a real Hilbert space, $r_i\in\mathcal{G}_i$, $A_i:\mathcal{G}_i\to 2^{\mathcal{G}_i}$ is maximal monotone, $D_i:\mathcal{G}_i\to 2^{\mathcal{G}_i}$ is maximal monotone and ν_i–strongly monotone for $\nu_i\in]0,+\infty[$, $A_i\Box D_i=(A_i^{-1}+D_i^{-1})^{-1}$ is the parallel sum of A_i and D_i, finally, let $L_i:\mathcal{H}\to\mathcal{G}_i$ be a non-zero bounded linear operator. Now, consider the following monotone inclusion problem,

$$\text{Find } x\in\mathcal{H} \text{ s.t. } 0\in Cx+Bx+\sum_{i=1}^n\omega_iL_i^*\big((A_i\Box D_i)(L_ix-r_i)\big), \tag{5.4}$$

and the corresponding dual problem,

$$\text{Find } (v_i \in \mathcal{G}_i, \cdots, v_n \in \mathcal{G}_n) \text{ s.t. } (\exists x \in \mathcal{H}), \begin{cases} 0 \in Cx + Bx + \sum_{i=1}^{n} \omega_i L_i^* v_i, \\ 0 \in (A_i \Box D_i)(L_i x - r_i) - v_i, \end{cases} \tag{5.5}$$

denote by $\mathcal{P}$ and $\mathcal{D}$ the solution sets of (5.4) and (5.5) respectively.

Algorithm 3: Vũ's Primal–Dual splitting algorithm

Input: Let $\tau, (\sigma_i)_i > 0$ such that $\eta = \min\{\frac{1}{\tau}, \frac{1}{\sigma_1}, ..., \frac{1}{\sigma_n}\}\big(1 - \sqrt{\tau \sum_i \sigma_i \omega_i \|L_i\|^2}\big)$, $2\eta\beta > 1$ where $\beta = \min\{\mu, \nu_1, ..., \nu_n\}$, and $\lambda_k \in]0, \frac{4\eta\beta-1}{2\eta\beta}]$. Let $\big(\varepsilon_1^k\big)_{k\in\mathbb{N}}$ and $\big(\varepsilon_2^k\big)_{k\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{H}$, $\big(\varepsilon_{3,i}^k\big)_{k\in\mathbb{N}}$ and $\big(\varepsilon_{4,i}^k\big)_{k\in\mathbb{N}}$ be absolutely summable sequences in $\mathcal{G}_i$ for $i \in [\![1, n]\!]$.

Initial: $k = 0$, $x^0 \in \mathcal{H}$, $v_i^0 \in \mathcal{G}_i$, $i \in [\![1, n]\!]$;

repeat

  $p^{k+1} = J_{\tau C}\big(x^k - \tau(\sum_i \omega_i L_i^* v_i^k + Bx^k + \varepsilon_1^k)\big) + \varepsilon_2^k$;

  $y^{k+1} = 2p^{k+1} - x^k$;

  $x^{k+1} = x^k + \lambda_k(p^{k+1} - x^k)$;

  for $i = 1$ **to** n **do**

    $q_i^{k+1} = J_{\sigma_i A_i^{-1}}\big(v_i^k + \sigma_i(L_i y_i^{k+1} - D_i^{-1} v_i^k - \varepsilon_{3,i}^k - r_i)\big) + \varepsilon_{4,i}^k$;

    $v_i^{k+1} = v_i^k + \lambda_k(q_i^{k+1} - v_i^k)$;

  $k = k + 1$;

until *convergence*;

Note that in finite–dimension, if $n = 1, r = 0, D = 0$ and let the iteration errorless, then if $\lambda_k \equiv 1$, the algorithm reduces to the method proposed in [14, Algorithm 1].

5.3.1 Fixed point formulation

In this subsection, we briefly recall the derivation of the fixed point iteration for Algorithm 3, more details can be found in [68, Section 3]. Define $\boldsymbol{\mathcal{G}} = \mathcal{G}_i \times \cdots \times \mathcal{G}_n$ be the real Hilbert space with the scalar product and the associated norm respectively defined as, for $\forall \boldsymbol{v}_1, \boldsymbol{v}_2 \in \boldsymbol{\mathcal{G}}$,

$$\langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle_{\boldsymbol{\mathcal{G}}} = \sum_{i=1}^{n} \omega_i \langle v_{1,i}, v_{2,i} \rangle_{\mathcal{G}_i}, \text{ and } \|\boldsymbol{v}_1\|_{\boldsymbol{\mathcal{G}}} = \sqrt{\textstyle\sum_{i=1}^{n} \omega_i \|v_{1,i}\|_{\mathcal{G}_i}^2}.$$

Next, let $\boldsymbol{\mathcal{K}} = \mathcal{H} \oplus \boldsymbol{\mathcal{G}}$, be the Hilbert direct sum, then the scalar product and norm of $\boldsymbol{\mathcal{K}}$ are respectively defined by

$$\langle (x_1, \boldsymbol{v}_1), (x_2, \boldsymbol{v}_2) \rangle_{\boldsymbol{\mathcal{K}}} = \langle x_1, x_2 \rangle + \langle \boldsymbol{v}_1, \boldsymbol{v}_2 \rangle_{\boldsymbol{\mathcal{G}}}, \ \|(x_1, \boldsymbol{v}_1)\|_{\boldsymbol{\mathcal{K}}} = \sqrt{\|x_1\|^2 + \|\boldsymbol{v}_1\|_{\boldsymbol{\mathcal{G}}}^2},$$

and define the following operators on $\boldsymbol{\mathcal{K}}$,

$$\begin{aligned}
\boldsymbol{C} &: \boldsymbol{\mathcal{K}} \to 2^{\boldsymbol{\mathcal{K}}}, (x, \boldsymbol{v}) \mapsto (Cx) \times \big(r_1 + A_1^{-1} v_1\big) \times \cdots \times \big(r_n + A_n^{-1} v_n\big), \\
\boldsymbol{D} &: \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}, (x, \boldsymbol{v}) \mapsto \big(\textstyle\sum_i \omega_i L_i^* v_i, -L_1 x, \cdots, -L_n x\big), \\
\boldsymbol{E} &: \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}, (x, \boldsymbol{v}) \mapsto \big(Bx, D_1^{-1} v_1, \cdots, D_n^{-1} v_n\big), \\
\boldsymbol{F} &: \boldsymbol{\mathcal{K}} \to \boldsymbol{\mathcal{K}}, (x, \boldsymbol{v}) \mapsto \Big(\tfrac{1}{\tau} x - \textstyle\sum_i \omega_i L_i^* v_i, \tfrac{1}{\sigma_1} v_1 - L_1 x, \cdots, \tfrac{1}{\sigma_n} v_n - L_n x\Big),
\end{aligned}$$

then it can be proved that $\boldsymbol{C}$ and $\boldsymbol{D}$ are maximal monotone, $\boldsymbol{E}$ is β–cocoercive, and $\boldsymbol{F}$ is self–adjoint and η–strongly positive.

We further define

$$\begin{aligned} s^{k+1} &= J_{\tau C}\big(x^k - \tau(\textstyle\sum_i \omega_i L_i^* v_i^k + Bx^k)\big), \\ t_i^{k+1} &= J_{\sigma_i A_i^{-1}}\big(v_i^k + \sigma_i(L_i y_i^{k+1} - D_i^{-1} v_i^k - r_i)\big), \end{aligned}$$

and the corresponding error

$$\begin{aligned} \varepsilon_s^k &= J_{\tau C}\big(x^k - \tau(\textstyle\sum_i \omega_i L_i^* v_i^k + Bx^k + \varepsilon_1^k)\big) - s^{k+1}, \\ \varepsilon_{t,i}^k &= J_{\sigma_i A_i^{-1}}\big(v_i^k + \sigma_i(L_i y_i^{k+1} - D_i^{-1} v_i^k - \varepsilon_{3,i}^k - r_i)\big) - t_i^{k+1}, \end{aligned}$$

Since $J_{\tau C}$ and $J_{\sigma_i A_i^{-1}}$ are firmly non-expansive, we have $\|\varepsilon_s^k\| \le \|\varepsilon_1^k\|$ and $\|\varepsilon_{t,i}^k\| \le \|\varepsilon_{3,i}^k\|$. Next denote

$$\begin{cases} \boldsymbol{u}^k = (p^k, q_1^k, \cdots, q_n^k), \quad \boldsymbol{z}^k = (x^k, v_1^k, \cdots, v_n^k), \\ \boldsymbol{t}^k = \big(s^k, t_1^k, ..., t_n^k\big), \quad \boldsymbol{\varepsilon}_{\boldsymbol{t}}^k = \big(\varepsilon_s^k + \varepsilon_2^k, \varepsilon_{t,1}^k + \varepsilon_{4,1}^k, ..., \varepsilon_{t,n}^k + \varepsilon_{4,n}^k\big), \\ \boldsymbol{\varepsilon}_1^k = (\varepsilon_2^k, \varepsilon_{4,1}^k, \cdots, \varepsilon_{4,n}^k), \quad \boldsymbol{\varepsilon}_3^k = (\varepsilon_1^k, \varepsilon_{3,1}^k, \cdots, \varepsilon_{3,n}^k), \\ \boldsymbol{\varepsilon}_4^k = \Big(\frac{1}{\tau}\varepsilon_2^k, \frac{1}{\sigma_1}\varepsilon_{4,1}^k, \cdots, \frac{1}{\sigma_n}\varepsilon_{4,n}^k\Big), \quad \boldsymbol{\varepsilon}_2^k = \boldsymbol{F}^{-1}\big((\boldsymbol{D}+\boldsymbol{E})\boldsymbol{\varepsilon}_1^k + \boldsymbol{\varepsilon}_3^k - \boldsymbol{\varepsilon}_4^k\big), \end{cases}$$

then there holds

$$\sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}_1^k\|_{\boldsymbol{\mathcal{K}}}, \ \sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}_3^k\|_{\boldsymbol{\mathcal{K}}}, \ \sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}_4^k\|_{\boldsymbol{\mathcal{K}}}, \ \sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}_2^k\|_{\boldsymbol{\mathcal{K}}}, \ \sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}_{\boldsymbol{t}}^k\|_{\boldsymbol{\mathcal{K}}} < +\infty. \tag{5.6}$$

Then the fixed point equation of Algorithm 3 is

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k\big(J_{\boldsymbol{A}}(\boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k - \boldsymbol{\varepsilon}_2^k) + \boldsymbol{\varepsilon}_1^k - \boldsymbol{z}^k\big), \tag{5.7}$$

where $\boldsymbol{A} = \boldsymbol{F}^{-1}(\boldsymbol{C}+\boldsymbol{D})$ and $\boldsymbol{B} = \boldsymbol{F}^{-1}\boldsymbol{E}$. Also $\boldsymbol{u}^{k+1} = J_{\boldsymbol{A}}(\boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k - \boldsymbol{\varepsilon}_2^k) + \boldsymbol{\varepsilon}_1^k$ and $\boldsymbol{t}^{k+1} = J_{\boldsymbol{A}}(\boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k)$. Now define the real Hilbert product space $\boldsymbol{\mathcal{K}}_{\boldsymbol{F}}$ with the scalar product and the associated norm defined by, $\forall \boldsymbol{z}_1, \boldsymbol{z}_2 \in \boldsymbol{\mathcal{K}}$,

$$\langle \boldsymbol{z}_1, \boldsymbol{z}_2\rangle_{\boldsymbol{F}} = \langle \boldsymbol{z}_1, \boldsymbol{F}\boldsymbol{z}_2\rangle_{\boldsymbol{\mathcal{K}}}, \text{ and } \|\boldsymbol{z}_1\|_{\boldsymbol{F}} = \sqrt{\langle \boldsymbol{z}_1, \boldsymbol{F}\boldsymbol{z}_1\rangle_{\boldsymbol{\mathcal{K}}}},$$

then it can be verified that $\boldsymbol{A}$ is maximal monotone on $\boldsymbol{\mathcal{K}}_{\boldsymbol{F}}$, $\boldsymbol{B}$ is $\eta\beta$–cocoercive on $\boldsymbol{\mathcal{K}}_{\boldsymbol{F}}$ with $2\eta\beta > 1$, this means Algorithm 3 is the FBS method under the metric $\boldsymbol{F}$. Define $\boldsymbol{T} = J_{\boldsymbol{A}} \circ (\mathbf{Id} - \boldsymbol{B})$, based on Proposition 5.1, we have

$$J_{\boldsymbol{A}} \in \mathcal{A}(\tfrac{1}{2}), \ \mathbf{Id} - \boldsymbol{B} \in \mathcal{A}(\tfrac{1}{2\eta\beta}) \text{ and } \boldsymbol{T} \in \mathcal{A}(\tfrac{2\eta\beta}{4\eta\beta-1}).$$

Let $\boldsymbol{\varepsilon}^k = \big(J_{\boldsymbol{A}}(\boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k - \boldsymbol{\varepsilon}_2^k) + \boldsymbol{\varepsilon}_1^k\big) - J_{\boldsymbol{A}}(\boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k)$, then from (5.6) we have $\sum_{k\in\mathbb{N}} \|\boldsymbol{\varepsilon}^k\| < +\infty$, and (5.7) is equivalent to

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k\big(\boldsymbol{T}\boldsymbol{z}^k + \boldsymbol{\varepsilon}^k - \boldsymbol{z}^k\big) = \boldsymbol{T}_k\boldsymbol{z}^k + \lambda_k\boldsymbol{\varepsilon}^k, \tag{5.8}$$

which is the same as (3.1). Hence, the solution set of Algorithm 3 is non-empty, furthermore, for $\forall \boldsymbol{z}^\star \in \mathrm{fix}\boldsymbol{T}$,

$$\begin{aligned} \boldsymbol{z}^\star \in \mathrm{fix}\boldsymbol{T} &\Longleftrightarrow \boldsymbol{z}^\star \in \mathrm{zer}(\boldsymbol{A}+\boldsymbol{B}) \\ &\Longleftrightarrow \boldsymbol{z}^\star \in \mathrm{zer}\boldsymbol{F}^{-1}(\boldsymbol{C}+\boldsymbol{D}+\boldsymbol{E}) \\ &\Longleftrightarrow \boldsymbol{z}^\star \in \mathrm{zer}(\boldsymbol{C}+\boldsymbol{D}+\boldsymbol{E}) \\ &\Longrightarrow x^\star \in \mathcal{P}, \ \boldsymbol{v}^\star \in \mathcal{D}. \end{aligned}$$

5.3.2 Iteration–complexity bounds

In the following, we present the pointwise and ergodic iteration–complexity bounds for Algorithm 3 under the product Hilbert space $\boldsymbol{\mathcal{K}}$. First we have

$$\boldsymbol{e}^k = \big(\mathbf{Id} - \boldsymbol{T}\big)\boldsymbol{z}^k = \boldsymbol{z}^k - \boldsymbol{t}^{k+1}.$$

Let $\underline{\tau} = \inf_{k\in\mathbb{N}} \lambda_k(\frac{4\eta\beta-1}{2\eta\beta} - \lambda_k)$, $\overline{\tau} = \sup_{k\in\mathbb{N}} \lambda_k(\frac{4\eta\beta-1}{2\eta\beta} - \lambda_k)$, $\delta = \max\{\frac{1}{\tau}, \frac{1}{\sigma_1}, ..., \frac{1}{\sigma_n}\}$. Define the distance $d_0 = \inf_{\boldsymbol{z}^\star\in\mathrm{fix}\boldsymbol{T}} \|\boldsymbol{z}^0 - \boldsymbol{z}^\star\|_{\boldsymbol{\mathcal{K}}}$, $\nu_1 = 2\sup_{k\in\mathbb{N}} \|\boldsymbol{T}_k\boldsymbol{z}^k - \boldsymbol{z}^\star\|_{\boldsymbol{\mathcal{K}}} + \sup_{k\in\mathbb{N}} \lambda_k\|\boldsymbol{\varepsilon}^k\|_{\boldsymbol{\mathcal{K}}}$ and $\nu_2 = 2\sup_{k\in\mathbb{N}} \|\boldsymbol{e}^k - \boldsymbol{e}^{k+1}\|_{\boldsymbol{\mathcal{K}}}$.

Define $\boldsymbol{v}^k = \boldsymbol{z}^k - \boldsymbol{z}^{k+1} + \lambda_k\boldsymbol{\varepsilon}^k$, $\boldsymbol{w}^{k+1} = \big(\boldsymbol{z}^{k+1} - (1-\lambda_k)\boldsymbol{z}^k\big)/\lambda_k - \boldsymbol{\varepsilon}^k$, $\boldsymbol{g}_{\boldsymbol{F}}^{k+1} = \boldsymbol{z}^k - \boldsymbol{B}\boldsymbol{z}^k - \boldsymbol{w}^{k+1}$.

Proposition 5.10 (Pointwise iteration–complexity bounds). *For the relaxed fixed point iteration (5.7), there holds*

$$\|\boldsymbol{e}^k\|_{\boldsymbol{\mathcal{K}}} \le \frac{2\delta}{\eta}\sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}} \quad \text{and} \quad \|\boldsymbol{v}^k\|_{\boldsymbol{\mathcal{K}}} \le \frac{2\delta\lambda_k}{\eta}\sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}},$$

and $\boldsymbol{g}_{\boldsymbol{F}}^{k+1} \in \boldsymbol{A}\boldsymbol{w}^{k+1}$,

$$d\Big(0, \big(\boldsymbol{A} + \boldsymbol{B}\big)\boldsymbol{w}^{k+1}\Big) \le \frac{2\delta}{\eta}\sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}},$$

where $C_1 = \nu_1 \sum_{j\in\mathbb{N}} \lambda_j\|\boldsymbol{\varepsilon}^j\|_{\boldsymbol{\mathcal{K}}} + \nu_2\overline{\tau}\sum_{\ell\in\mathbb{N}}(\ell+1)\|\boldsymbol{\varepsilon}^\ell\|_{\boldsymbol{\mathcal{K}}} < +\infty$.

Remark 5.11. Since $\boldsymbol{z}^\star \in \mathrm{fix}(\boldsymbol{A} + \boldsymbol{B}) \Leftrightarrow \boldsymbol{z}^\star \in \mathrm{zer}\boldsymbol{F}^{-1}(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{E})$, Proposition 5.10 means that for the monotone inclusion $0 \in \boldsymbol{F}^{-1}(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{E})\boldsymbol{z}$, we have

$$d\Big(0, \big(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{E}\big)\boldsymbol{w}^{k+1}\Big) \le \frac{2\delta}{\eta}\sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}},$$

namely, the method can find an ϵ–accurate solution for $0 \in \big(\boldsymbol{C} + \boldsymbol{D} + \boldsymbol{E}\big)\boldsymbol{z}$ in at most $O(1/\epsilon)$ iterations.

Let $\bar{\boldsymbol{v}}^k = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{v}^j$, $\bar{\boldsymbol{z}}^k = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{z}^j$, $\bar{\boldsymbol{w}}^{k+1} = \frac{1}{k+1}\sum_{j=0}^k \boldsymbol{w}^j$, $\bar{\boldsymbol{g}}_{\boldsymbol{F}}^{k+1} = \frac{1}{\gamma}\bar{\boldsymbol{z}}^k - \boldsymbol{B}\bar{\boldsymbol{z}}^k - \frac{1}{\gamma}\bar{\boldsymbol{w}}^{k+1}$.

Proposition 5.12 (Ergodic iteration–complexity bounds). *The following statements hold, if* $C_2 = \sum_{k\in\mathbb{N}} \lambda_k\|\boldsymbol{\varepsilon}^k\|_{\boldsymbol{\mathcal{K}}} < +\infty$,

$$\|\bar{\boldsymbol{e}}^k\|_{\boldsymbol{\mathcal{K}}} \le \frac{4\delta(d_0 + C_2)}{\eta\Lambda_k}, \ \|\bar{\boldsymbol{v}}^k\|_{\boldsymbol{\mathcal{K}}} \le \frac{4\delta(d_0 + C_2)}{\eta(k+1)},$$

and let $\underline{\lambda} = \inf_{k\in\mathbb{N}} \lambda_k$,

$$\|\bar{\boldsymbol{g}}_{\boldsymbol{F}}^{k+1} + \boldsymbol{B}\bar{\boldsymbol{w}}^{k+1}\|_{\boldsymbol{\mathcal{K}}} \le \frac{4\delta(d_0 + C_2)}{\eta\gamma\underline{\lambda}(k+1)}.$$

Furthermore, if we reformulated (5.5) to the following format,

$$\text{Find } \boldsymbol{v} \in \boldsymbol{\mathcal{G}} \text{ s.t. } (\exists x \in \mathcal{H}), \begin{cases} 0 \in Cx + Bx + \sum_{i=1}^n \omega_i L_i^* v_i, \\ 0 \in \big(A_i^{-1} + D_i^{-1}\big)v_i - (L_i x - r_i), i \in [\![1, n]\!], \end{cases} \tag{5.9}$$

then from Algorithm 3, we have

$$\big(\tfrac{1}{\tau}\mathrm{Id}-B\big)\,x^k-\big(\tfrac{1}{\tau}\mathrm{Id}-B\big)\,s^{k+1}\in(C+B)s^{k+1}+\sum_{i=1}^{n}\omega_i L_i^* v_i^k,$$

and for every $i\in[\![1,n]\!]$,

$$\Big(\tfrac{1}{\sigma_i}\mathrm{Id}-D_i^{-1}\Big)\,v_i^k-\Big(\tfrac{1}{\sigma_i}\mathrm{Id}-D_i^{-1}\Big)\,t_i^{k+1}\in\big(A_i^{-1}+D_i^{-1}\big)t_i^{k+1}-\big(L_iy^{k+1}-r_i\big).$$

Define operator,

$$\boldsymbol{G}:\ \boldsymbol{\mathcal{K}}\to 2^{\boldsymbol{\mathcal{K}}},\ (x,\boldsymbol{v})\mapsto\big((\tfrac{\mathrm{Id}}{\tau}-B)x\big)\times\big((\tfrac{\mathrm{Id}}{\sigma_1}-D_1^{-1})v_1\big)\times\cdots\times\big((\tfrac{\mathrm{Id}}{\sigma_n}-D_n^{-1})v_n\big),$$

then we have

$$\boldsymbol{g}^{k+1}=\boldsymbol{G}\boldsymbol{z}^k-\boldsymbol{G}\boldsymbol{t}^{k+1}\in\begin{pmatrix}(C+B)s^{k+1}+\sum_i\omega_iL_i^*v_i^k\\ (A_1^{-1}+D_1^{-1})t_1^{k+1}-\big(L_1y^{k+1}-r_1\big)\\ \vdots\\ (A_n^{-1}+D_n^{-1})t_n^{k+1}-\big(L_ny^{k+1}-r_n\big)\end{pmatrix}.$$

Denote the right hand side term of the inclusion as $\boldsymbol{M}\big(\boldsymbol{v}^k,\boldsymbol{t}^{k+1},y^{k+1}\big)$

Proposition 5.13 (Iteration–complexity bound for Dual inclusion (5.9)). *For the Algorithm 3, there holds*

$$d\Big(0,\boldsymbol{M}\big(\boldsymbol{v}^k,\boldsymbol{t}^{k+1},y^{k+1}\big)\Big)\le\frac{2\delta^2}{\eta}\sqrt{\frac{d_0^2+C_1}{\underline{\tau}(k+1)}}.$$

Remark 5.14. Criterion $\|\boldsymbol{g}^{k+1}\|_{\boldsymbol{\mathcal{K}}}$ demonstrates that the algorithm can find an ϵ–solution of (5.5), *i.e.*, a pair $(\boldsymbol{z},\boldsymbol{u})$ such that

$$\left\|\begin{matrix}(C+B)p+\sum_{i=1}^n\omega_iL_i^*v_i\\ (A_1^{-1}+D_1^{-1})q_1-\big(L_1(2p-x)-r_1\big)\\ \vdots\\ (A_n^{-1}+D_n^{-1})q_n-\big(L_n(2p-x)-r_n\big)\end{matrix}\right\|_{\boldsymbol{\mathcal{K}}}^2\le\epsilon$$

in at most $O(1/\epsilon)$ iterations.

6 Non-stationary Krasnosel'skiĭ–Mann iteration

6.1 General case

The fixed point iteration discussed in Section 3 is stationary, that is, operator T is fixed during the iterations. In this section, we study the non-stationary case of (3.1), and show that, under mild assumptions, the non-stationary case can be seen as a perturbation of the stationary one. In order to ensure convergence, the perturbation error should be absolutely summable.

For the methods discussed in Section 5, FBS/GFB and DRS methods for instance, their fixed point operators are characterised by a parameter γ, which is a constant along the iterations. For the FBS method, as stated in [21, Theorem 3.4], if $\lambda_k\in]0,1[$, $(\gamma_k)_{k\in\mathbb{N}}$ can be varying in $]0,2\beta[$. In the following context, we first investigate the convergence and iteration–complexity bounds of the non-stationary version of (3.1), and then specialize the result to the GFB method.

Let $T_\Gamma : \mathcal{H} \to \mathcal{H}$ be a family of non-expansive operators depending on a parameter Γ, $\lambda_k \in]0,1[$, and the non-stationary fixed point iteration is defined by

$$\begin{aligned} z^{k+1} &= z^k + \lambda_k(T_{\Gamma_k} z^k + \varepsilon^k - z^k) \\ &= \big(\lambda_k T_{\Gamma_k} + (1-\lambda_k)\mathrm{Id}\big) z^k + \lambda_k \varepsilon^k \\ &= T_{\Gamma_k,\lambda_k} z^k + \lambda_k \varepsilon^k. \end{aligned} \tag{6.1}$$

If we define $\varepsilon^k_{\Gamma_k} = (T_{\Gamma_k} - T_\Gamma) z^k$, $\pi^k = \varepsilon^k_{\Gamma_k} + \varepsilon^k$, then (6.1) can be simplified to

$$z^{k+1} = \big(\lambda_k T_\Gamma + (1-\lambda_k)\mathrm{Id}\big) z^k + \lambda_k \pi^k = T_{\Gamma,\lambda_k} z^k + \lambda_k \pi^k, \tag{6.2}$$

and the corresponding e^k of (6.1) is

$$e^k = \frac{z^k - z^{k+1}}{\lambda_k} + \pi^k.$$

When comparing this iteration (6.2) to Definition 3.1, a new error term π is introduced. To obtain the convergence of the non-stationary iteration, we adapt the work from different contexts: [39, Proposition 2.1] (for Banach spaces endowed with an appropriate compatible topology) and [40, Proposition 3.1] (for real Hilbert spaces).

Theorem 6.1 (Convergence of iteration (6.1)). *If the following assumptions*

(a) $\mathrm{fix} T_\Gamma \neq \emptyset$;

(b) *for* $\forall k \in \mathbb{N}$, T_{Γ_k,λ_k} *is* $(1+\beta_k)$*–Lipschitz with* $\beta_k \geq 0$, *and* $(\beta_k)_{k\in\mathbb{N}} \in \ell^1_+$;

(c) $\lambda_k \in]0,1[$ *such that* $\underline{\tau} > 0$, *and the property is translation invariant;*

(d) $\big(\lambda_k \|\varepsilon^k\|\big) \in \ell^1_+$;

(e) *for* $\forall \rho \in [0,+\infty[$, *sequence* $(\lambda_k \Delta_{k,\rho})_{k\in\mathbb{N}}$ *is summable, where*

$$\Delta_{k,\rho} = \sup_{\|z\|\leq\rho} \|T_{\Gamma_k,\lambda} z - T_{\Gamma,\lambda} z\|,$$

are satisfied, then $(e^k)_{k\in\mathbb{N}}$ *converges strongly to* 0, $(z^k)_{k\in\mathbb{N}}$ *converges weakly to a point* $z^\star \in \mathrm{fix} T_\Gamma$.

This theorem indicates that the perturbed approximate method can be seen as an approximate version of the exact method with an extra error term which should also be summable owing to assumption (e) of Theorem 6.1.

Next, we discuss one specific scenario that $\forall k \in \mathbb{N}$, T_{Γ_k,λ_k} is α_k–averaged. By definition, for $\forall k \in \mathbb{N}$, there exists a 1–Lipschitz operator $R_{\Gamma_k,\lambda_k} : \mathcal{H} \to \mathcal{H}$ such that $T_{\Gamma_k,\lambda_k} = \alpha R_{\Gamma_k,\lambda_k} + (1 - \alpha_k)\mathrm{Id}$. Defining $\lambda'_k = \alpha_k \lambda_k$, then we have the following corollary.

Corollary 6.2. *Assume (a), (d) of Theorem 6.1 hold and if the following assumptions*

(a) *for* $\forall k \in \mathbb{N}$, T_{Γ_k,λ_k} *is* α_k*–averaged,* $\alpha_k \in]0,1[$;

(b) $\lambda_k \in]0, \frac{1}{\alpha_k}[$, *such that* $\inf_{k\in\mathbb{N}} \lambda'_k(1-\lambda'_k) > 0$, $\sum_{k\in\mathbb{N}} \lambda'_k(1-\lambda'_k) = +\infty$;

(c) *for* $\forall \rho \in [0,+\infty[$, *sequence* $(\lambda'_k \Delta_{k,\rho})_{k\in\mathbb{N}}$ *is summable, where*

$$\Delta_{k,\rho} = \sup_{\|z\|\leq\rho} \|R_{\Gamma_k,\lambda} z - T_{\Gamma,\lambda} z\|,$$

are satisfied, then $(e^k)_{k\in\mathbb{N}}$ *converges strongly to* 0*,* $(z^k)_{k\in\mathbb{N}}$ *converges weakly to a point* $z^\star \in \mathrm{fix}T_\Gamma$*.*

Assumptions (d), (e) of Theorem 6.1 imply that

$$\sum_{k\in\mathbb{N}} \lambda_k\|\pi^k\| \le \sum_{k\in\mathbb{N}} \lambda_k\big(\|\varepsilon^k_{\Gamma_k}\| + \|\varepsilon^k\|\big) < +\infty,$$

therefore, if we can further impose an assumption on $(\pi^k)_{k\in\mathbb{N}}$ as in Theorem 3.3, then we can obtain the iteration–complexity bounds for the non-stationary iteration (6.1). Define $\Lambda_k = \sum_{j=0}^k \lambda_j$, $\bar{e}^k = \frac{1}{\Lambda_k}\sum_{j=0}^k \lambda_j e^j$, and let d_0 be the distance from the starting point z^0 to the fixed point set $\mathrm{fix}T_\Gamma$.

Theorem 6.3. *If the following assumptions*

(a) $\mathrm{fix}T_\Gamma \neq \emptyset$*;*

(b) *for* $\forall k \in \mathbb{N}$*,* T_{Γ_k,λ_k} *is non-expansive;*

(c) $\lambda_k \in]0,1[$*, such that* $\inf_{k\in\mathbb{N}} \lambda_k(1-\lambda_k) > 0$*;*

(d) $\sum_{k\in\mathbb{N}}(k+1)\|\pi^k\| < +\infty$*;*

are hold, then we have

(i) *Pointwise iteration–complexity bound:*

$$\|e^k\| \le \sqrt{\frac{d_0^2 + C_1'}{\underline{\tau}(k+1)}};$$

where $C_1' = \nu_1\sum_{j\in\mathbb{N}}\lambda_j\|\pi^j\| + \nu_2\overline{\tau}\sum_{\ell\in\mathbb{N}}(\ell+1)\|\pi^\ell\| < +\infty$*.*

(ii) *Ergodic iteration–complexity bound: let* $C_2' = \sum_{j\in\mathbb{N}}\lambda_j\|\pi^j\| < +\infty$*, then,*

$$\|\bar{e}^k\| \le \frac{2(d_0 + C_2')}{\Lambda_k}.$$

If $\inf_{k\in\mathbb{N}}\lambda_k > 0$*, we get* $O(1/k)$ *ergodic iteration–complexity bound for* (6.1)*.*

Example: Non-stationary GFB

As discussed in Subsection 5.1, the fixed point operator $\boldsymbol{T}_\gamma$ of GFB depends on a parameter γ. Now let γ varies during the iteration, and we have the following corresponding operators

$$\boldsymbol{T}_{1,\gamma_k} = \tfrac{1}{2}\big(R_{\gamma_k\boldsymbol{A}}R_{\boldsymbol{\mathcal{S}}} + \mathbf{Id}\big),\ \boldsymbol{T}_{2,\gamma_k} = \mathbf{Id} - \gamma_k\boldsymbol{B}_{\boldsymbol{\mathcal{S}}},\ \text{and}\ \boldsymbol{T}_{\gamma_k} = \boldsymbol{T}_{1,\gamma_k}\circ\boldsymbol{T}_{2,\gamma_k},$$

for $(\gamma_k)_{k\in\mathbb{N}} \in]0,2\beta[$, $\boldsymbol{T}_{\gamma_k}$ is $(\frac{2\beta}{4\beta-\gamma_k})$–averaged. Let $\boldsymbol{\varepsilon}^k_{\gamma_k} = (\boldsymbol{T}_{\gamma_k} - \boldsymbol{T}_\gamma)\boldsymbol{z}^k$, $\boldsymbol{\pi}^k = \boldsymbol{\varepsilon}^k_{\gamma_k} + \boldsymbol{\varepsilon}^k$, then the non-stationary version of (5.2) is defined by

$$\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k\big(\boldsymbol{T}_\gamma\boldsymbol{z}^k + \boldsymbol{\pi}^k - \boldsymbol{z}^k\big). \tag{6.3}$$

For $\gamma \in]0,2\beta[$, there exists a non-expansive operator $\boldsymbol{R}_\gamma : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ such that $\boldsymbol{T}_\gamma = \alpha\boldsymbol{R}_\gamma + (1-\alpha)\mathbf{Id}$, and for $\forall k \in \mathbb{N}$, there exists a non-expansive operator $\boldsymbol{R}_{\gamma_k} : \boldsymbol{\mathcal{H}} \to \boldsymbol{\mathcal{H}}$ such that $\boldsymbol{T}_{\gamma_k} = \alpha\boldsymbol{R}_{\gamma_k} + (1-\alpha_k)\mathbf{Id}$, define $\lambda_k' = \alpha_k\lambda_k$.

Theorem 6.4. *For the non-stationary iteration* (6.3)*, if the following assumptions hold*

(a) $\operatorname{zer}\big(B + \sum_i A_i\big) \neq \emptyset$;

(b) $\lambda_k \in]0, \frac{1}{\alpha_k}[$, *such that* $\inf_{k\in\mathbb{N}} \lambda_k(\frac{1}{\alpha_k} - \lambda_k) > 0$;

(c) $\big(\lambda_k \|\boldsymbol{\varepsilon}^k\|\big) \in \ell_+^1$;

(d) *for* $\forall \rho \in [0, +\infty[$, *sequence* $(\alpha_k \Delta_{k,\rho})_{k\in\mathbb{N}}$ *is summable, where*

$$\Delta_{k,\rho} = \sup_{\|\boldsymbol{z}\|\leq\rho} \|\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \boldsymbol{R}_\gamma \boldsymbol{z}\|,$$

(e) $(\gamma_k)_{k\in\mathbb{N}} \in]0, 2\beta[$ *such that* $0 < \underline{\gamma} \leq \gamma_k \leq \overline{\gamma} < 2\beta$, $\gamma \in [\underline{\gamma}, \overline{\gamma}]$, *and* $(|\gamma_k - \gamma|)_{k\in\mathbb{N}}$ *is summable;*

then, the sequence $(\boldsymbol{z}^k)_{k\in\mathbb{N}}$ *generated by* (6.3) *converges weakly to a point in* $\operatorname{fix}\boldsymbol{T}_\gamma$.

If we further assume that $\sum_{k\in\mathbb{N}}(k+1)\|\boldsymbol{\pi}^k\| < +\infty$, *then we obtain the iteration–complexity bounds for the non-stationary version of GFB algorithm as stated in Theorem 6.3.*

Remark 6.5. Corollary 6.4 is also applicable to the Douglas–Rachoford splitting method and the FDRS method [7].

7 Numerical experiments

To demonstrate the established iteration–complexity bounds and local convergence rate. In this section, we take 3 inverse problems as example: 1) anisotropic total variation (TV) based deconvolution with box constraint, 2) matrix completion with non-negativity constraints (NMC), 3) principal component pursuit problem (PCP) with application to video background and foreground decomposition. All the problems are solved by both GFB and Vũ's algorithm (vPDS).

7.1 Anisotropic TV deconvolution

Suppose the blurred observation $y \in \mathbb{R}^n$ is the convolution of $x_0 \in \mathbb{R}^n$ and a point spread function (PSF) h contanminated with additive white Gaussian noise w, which reads

$$y = \mathcal{M}(x_0) + w,$$

where $\mathcal{M} : \mathbb{R}^n \to \mathbb{R}^n$ is the linear operator associated to h. The deconvolution procedure is to provably recover or approximate x_0 from y, here we consider the anisotropic TV [61] based deconvolution model which is

$$\min_{x\in\mathbb{R}^{m\times n}} \frac{1}{2}\|\mathcal{M}(x) - y\|^2 + \mu\|\nabla x\|_1 + \iota_\Omega(x), \tag{7.1}$$

where $\mu > 0$ is the regularization parameter determined based on the noise level, $\iota_\Omega(\cdot)$ is the indicator funciton of box constraint, for instance $\Omega = [0, 255]^n$ if x is gray scale image. The problem can be solved by both GFB and vPDS methods, where the proximity operator of $\iota_\Omega(\cdot)$ is projection onto Ω, and for GFB method, the proximity operator of $\|\nabla \cdot\|_1$ is computed by minimum graph–cut [13, 37].

Figure 3 displays the observed pointwise and ergodic rates of $\|e^k\|$ and the theoretical bounds given by Theorem 3.3 and 3.6. Pointwise convergence rate is shown in subfigure (a) and (c), whose left half is log-log plot while the right half is semilog plot. As predicted by Theorem 3.3, globally $\|e^k\|$ converges at the rate of $O\big(1/\sqrt{k}\big)$. Then for a sufficiently large iteration number, a linear convergence regime takes over as clearly seen from the semilog plot, which is in consistent with the result of Theorem 4.2. Let us mention that the local linear convergence curve (dashed line) is fitted to the observed one, since the regularity modulus necessary to compute the theoretical rate in Theorem 4.2 is not easy to estimate. For the ergodic convergence, subfigure (b) and (d) of Figure 3, $O(1/k)$ convergence rates are observed which coincides with Theorem 3.6.

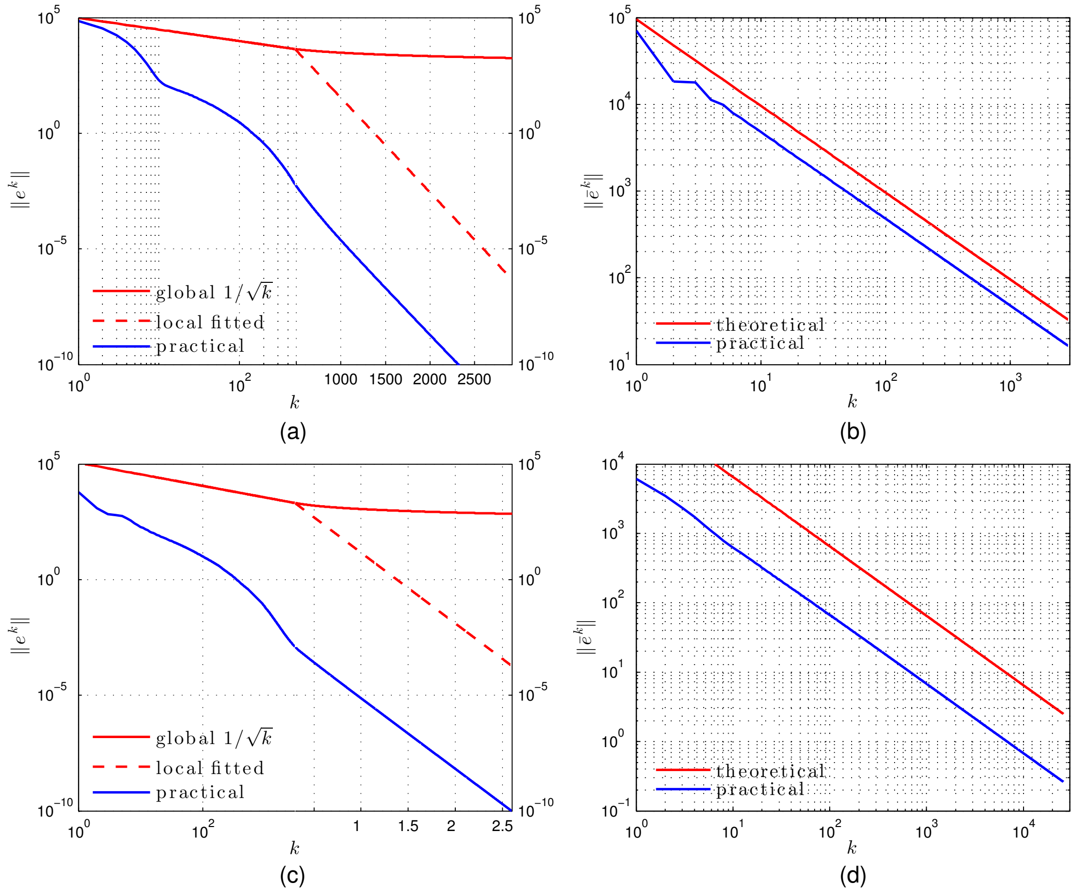


Figure 3: TV deconvolution of the cameraman image. (a) pointwise convergence of GFB, (b) ergodic convergence of GFB, (c) pointwise convergence of vPDS, (d) ergodic convergence of vPDS. Both methods achieve local linear convergence. Note that $\big(\|e^k\|\big)_{k\in\mathbb{N}}$ is non-increasing, which coincides with Lemma A.5 when $\varepsilon^k = 0$.

7.2 Non-negative matrix completion

Suppose we observe measurements $y \in \mathbb{R}^p$ of a low rank matrix $X_0 \in \mathbb{R}^{m\times n}$ with non-negative entries

$$y = \mathcal{M}(X_0) + w,$$

where $\mathcal{M} : \mathbb{R}^{m\times n} \to \mathbb{R}^p$ is a measurement operator, and w is the noise. In our experiment here, $\mathcal{M}$ selects p entries of its argument uniformly at random. The matrix completion problem consists in recovering X_0, or finding an approximation of it, by solving a convex optimization problem, namely the minimization of the nuclear norm [11, 12, 58]. In penalized form, the problem reads

$$\min_{X\in\mathbb{R}^{m\times n}} \frac{1}{2}\|y - \mathcal{M}(X)\|^2 + \mu\|X\|_* + \iota_{P_+}(X), \tag{7.2}$$

where $\iota_{P_+}(\cdot)$ is the indicator function of the non-negative orthant accounting for the non-negativity constraint, and $\mu > 0$ is a regularization parameter typically chosen proportional to the noise level. The proximity operator of both $\|\cdot\|_*$, $\iota_{P_+}(\cdot)$ have explicit forms, since $\mathrm{prox}_{\|\cdot\|_*}(X)$ amounts to soft–thresholding the singular values of X and $\mathrm{prox}_{\iota_{P_+}}(X)$ is the projector on the non-negative orthant.

Figure 4 displays the observed pointwise and ergodic rates of $\|e^k\|$ and the theoretical bounds computed given by Theorem 3.3 and 3.6. Both global and local convergence behaviours are similar to those observed in Figure 3.

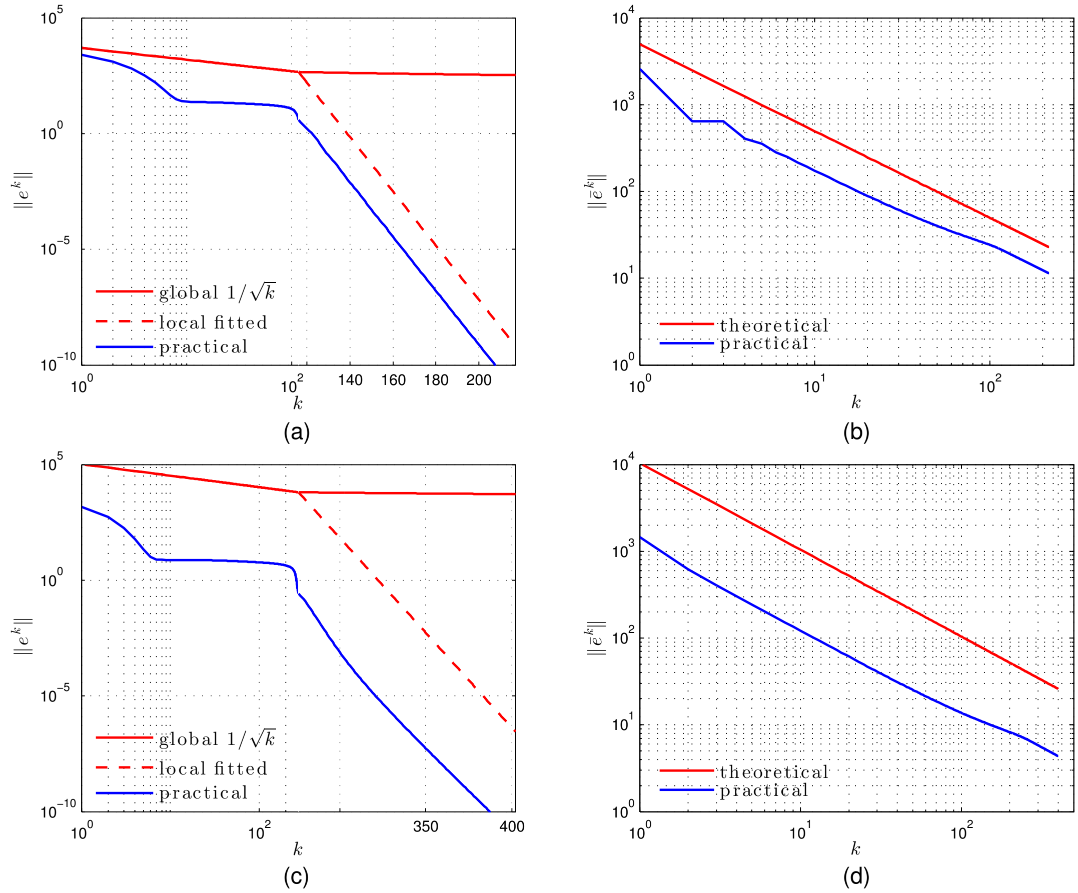


Figure 4: The size of the matrix X is 400×300, $\mathrm{rank}(X) = 20$ and the operator $\mathcal{M}$ is random projection mask. (a) pointwise convergence of GFB, (b) ergodic convergence of GFB, (c) pointwise convergence of vPDS, (d) ergodic convergence of vPDS.

7.3 Principal component pursuit

In this experiment, we consider the PCP problem [10], and apply it to decompose a video sequence into its background and foreground components. The rationale behind this is that since the background is virtually the same in all frames, if the latter are stacked as columns of a matrix, it is likely to be low–rank (even of rank 1 for perfectly constant background). On the other hand, moving objects appear occasionally on each frame and occupy only a small fraction of it. Thus the corresponding component would be sparse.

Assume that a real matrix $M \in \mathbb{R}^{m\times n}$ can be written as

$$M = X_{L,0} + X_{S,0} + w,$$

where $X_{L,0}$ is low–rank, $X_{S,0}$ is sparse and w is a perturbation matrix of variance σ that accounts for model imperfection (noise). The PCP proposed in [10] attempts to provably recover $(X_{L,0}, X_{S,0})$ to a good approximation, by solving a convex optimization. Here, toward an application to video decomposition, we also add a non-negativity constraint to the low–rank component, which leads to the following convex problem

$$\min_{X_L, X_S \in \mathbb{R}^{m\times n}} \frac{1}{2}\|M - X_L - X_S\|_F^2 + \mu_1 \|X_S\|_1 + \mu_2 \|X_L\|_* + \iota_{P_+}(X_L), \tag{7.3}$$

where $\|\cdot\|_F$ is the Frobenius norm.

Observe that for fixed X_L, the minimizer of (7.3) is $X_S^\star = \mathrm{prox}_{\mu_1\|\cdot\|_1}(M - X_L)$. Thus, (7.3) is equivalent to

$$\min_{X_L} {}^1\big(\mu_1\|\cdot\|_1\big)(M - X_L) + \mu_2\|X_L\|_* + \iota_{P_+}(X_L), \tag{7.4}$$

where ${}^1\big(\mu_1\|\cdot\|_1\big)(M - X_L) = \min_Z \frac{1}{2}\|M - X_L - Z\|_F^2 + \mu_1\|Z\|_1$ is the Moreau Envelope of $\mu_1\|\cdot\|_1$ of index 1, and hence has 1–Lipschitz continuous gradient.

We first use a synthetic example to demonstrate the comparison of the two methods, as shown in Figure 5. Pointwise convergence rate of $\|e^k\|$ is shown in subfigure (a) and (c). Then in subfigure (b) and (d), we display the convergence behaviour of the criteria provided in Proposition 5.5 and 5.13.

Campus view video The video sequence consists of 400 frames, each of resolution 288×384 stacked as a column of the matrix M. Hence M is of size 110592×400. We then solved (7.4) to decompose the video into its foreground and background.

For the video test, we choose GFB method to solve the problem since it's faster than vPDS. Figure 6 displays the observed pointwise and ergodic rates and those predicted by Proposition 5.3–5.6. Figure 7 shows the decomposition of the video sequence, column (a) shows 3 frames from the video, column (b) and (c) are the corresponding low–rank component X_L and sparse component X_S. Notice that X_L correctly recovers the background, while X_S correctly identifies the moving pedestrians and their shadows.

7.4 Comparison of different choices of the relaxation parameter λ_k

In this part, we discuss the differences of the choice of γ and λ_k on the convergence of the error sequence $\|e^k\|$. The comparison is conducted via the PCP problem and GFB method.

From Proposition 2.4 (ii) and Remark 2.5, we have two bounds for the averaged modulus of the fixed point operator $\boldsymbol{T}$ of GFB, for $\lambda_k \in]0, \frac{1}{\alpha}[$, clearly $\alpha = \frac{2\beta}{4\beta - \gamma}$ provides a larger range for λ_k. And the following comparison shows that for appropriate choice of $\gamma \in]0, 2\beta[$, bigger λ_k leads to faster convergence.

For a given $\gamma \in]0, 2\beta[$, we have 8 different λ_k's, among which, 6 are constant, one is uniformly random in $]0, \frac{1}{\alpha}[$ and marked as "rnd $\lambda \in]0, 1/\alpha[$", and the other one is non-linear monotone increasing in interval $]\frac{1}{2\alpha}, \frac{1}{\alpha}[$ and marked as "inc $\lambda \in]1/(2\alpha), 1/\alpha[$". $\alpha = \frac{2\beta}{4\beta-\gamma}$ is chosen, and the comparison is demonstrated in Figure 8. The plot is split into two parts, the first 100 steps of error are plotted in log-log scale, while the rests are plotted by semilog scale.

Comparing the log-log part of the errors, it can be observed that,

(i) At the first step $k = 1$, $\|e^k\|$ obtains the deepest decay, for different γ's, it happens when $\lambda_k = 1$;

(ii) The bigger $|\lambda_k - 1|$ is, the slower $\|e^k\|$ decreases;

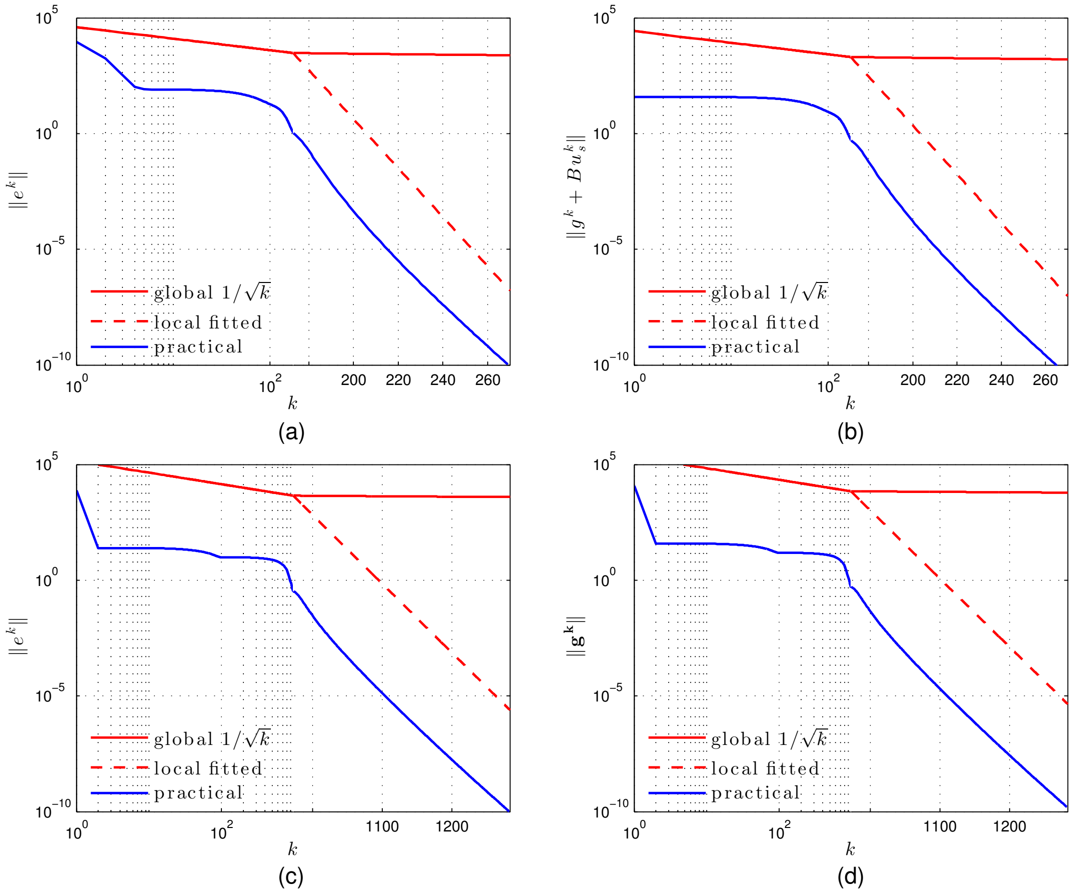


Figure 5: The size of the matrix X is 400×300, $\text{rank}(X) = 20$ and the operator $\mathcal{M}$ is random projection mask. (a) pointwise convergence of $\|e^k\|$ of GFB, (b) pointwise convergence of $\|g^k + B(\sum_i \omega_i u_i^k)\|$ of GFB, (c) pointwise convergence of $\|e^k\|$ of vPDS, (d) pointwise convergence $\|\boldsymbol{g}^k\|$ of vPDS.

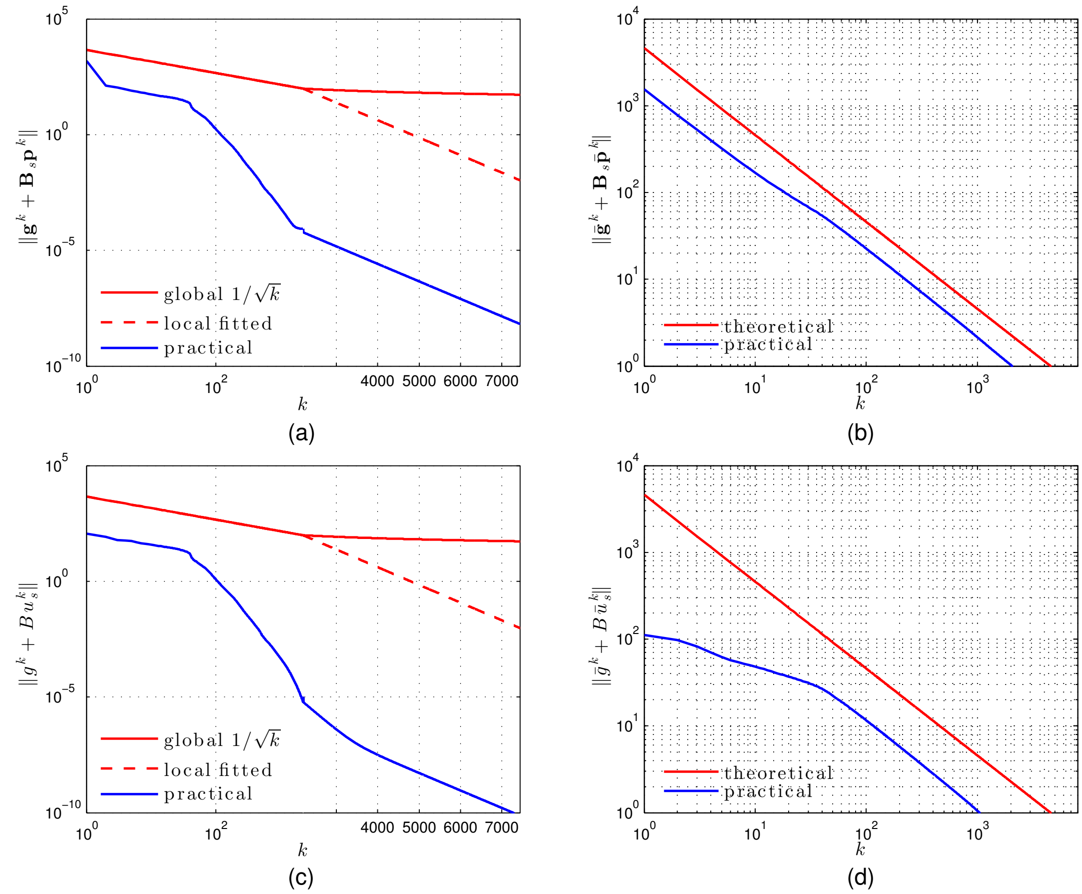
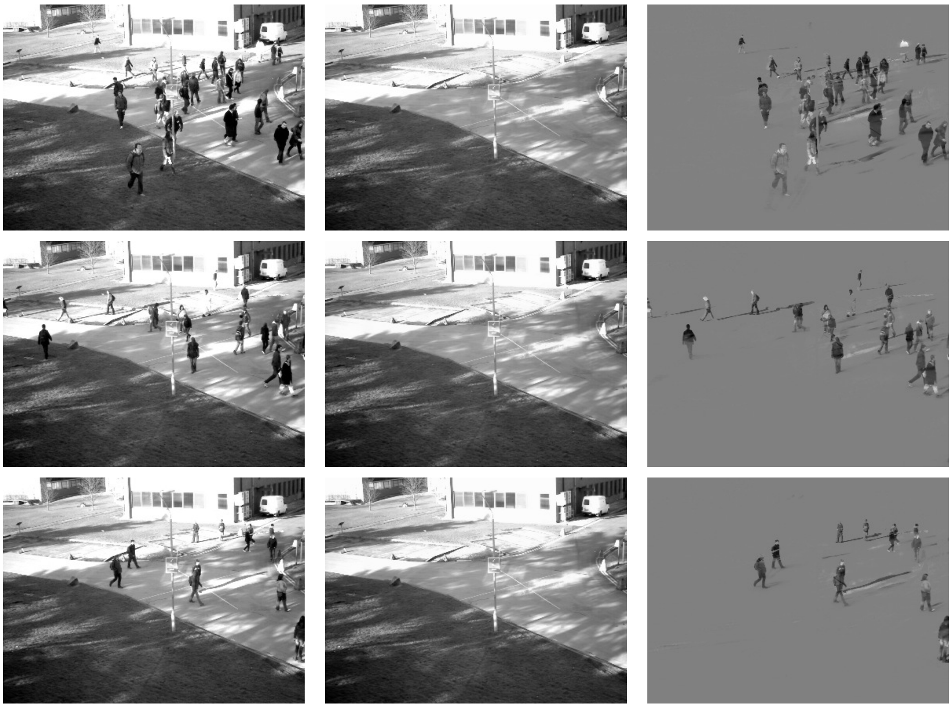


Figure 6: Observed rates and theoretical bounds of GFB. (a) pointwise convergence of $\|\boldsymbol{g}^k + \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}(\boldsymbol{p}^k)\|$, (b) ergodic convergence of $\|\bar{\boldsymbol{g}}^k + \boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\bar{\boldsymbol{p}}^k\|$, (c) pointwise convergence of $\|g^k + B(\sum_i \omega_i u_i^k)\|$, (d) ergodic convergence of $\|\bar{g}^k + B(\sum_i \omega_i \bar{u}_i^k)\|$.



(a) Original video M (b) Low rank X_L (c) Sparse X_S

Figure 7: Three frames from a 400–frame sequence taken in campus. (a) Original video M, (b) Low rank X_L and (c) Sparse X_S obtained by (7.3).

(iii) The speed of the two non-constant λ_k's depends on its values and no better than the constant ones.

Then for the semiolog part of the plot, we have the following observation,

(i) for the smaller γ's, $\lambda_k = \frac{1}{1.05\alpha}$ is the fastest linear rate;

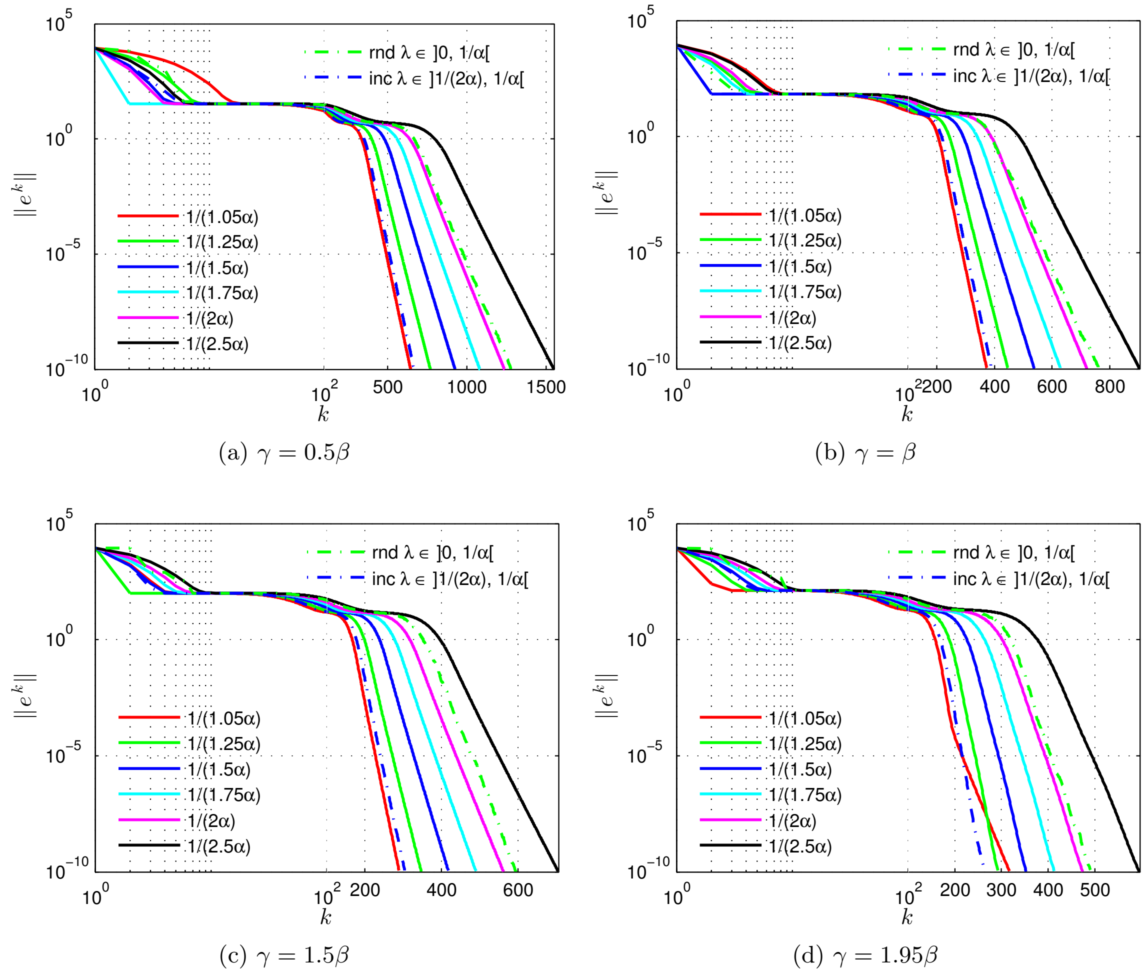
(ii) the random λ_k coincides with $\lambda_k = \frac{1}{2\alpha}$, with the fact that mean$(\lambda_k) = \frac{1}{2\alpha}$;

(iii) for $\lambda_k \in]\frac{1}{2\alpha}, \frac{1}{\alpha}[$ and non-decreasing, it has the same speed as $\lambda_k = \frac{1}{1.05\alpha}$.

For the theoretical bounds in Section 3, Corollary 3.5 and Theorem 4.2, the results imply that when $\lambda_k = \frac{1}{2\alpha}$, theoretical bounds obtain the best constant, while the practical error sequence shows that, for a given α, $\lambda \approx \frac{1}{\alpha}$ obtains the fastest convergence speed.

7.5 Non-stationary iteration

Now we illustrate the non-stationary iteration of GFB applied to PCP problem. The above comparison indicates that in practice we can choose the relaxation parameter λ_k as $\lambda_1 = 1$ and $\lambda_k \approx \frac{1}{\alpha}$, $\forall k \geq 2$. Next, we compare this setting with the non-stationary GFB, for the stationary case, we let $\gamma = 1.5\beta, \lambda_k = \{1, \frac{1}{1.05\alpha}, \cdots\}$, for the non-stationary case, let $\gamma_0 = 1.5\beta$, then 2 scenarios of γ_k are considered, $\gamma_{1,k} = \gamma_0 + \frac{1.9\beta-\gamma_0}{1.1^k}$, $\gamma_{2,k} = \gamma_0 + \frac{1.9\beta-\gamma_0}{k^2}$. The result is given in Figure 9. Note that for both varying γ_k's, $(|\gamma_k - 1.5\beta|)_{k\in\mathbb{N}} \in \ell^1_+$, however, much more iterations are required by $(\gamma_{2,k})_{k\in\mathbb{N}}$ as it converges slower to 1.5β than $(\gamma_{1,k})_{k\in\mathbb{N}}$.



(a) $\gamma = 0.5\beta$ (b) $\gamma = \beta$

(c) $\gamma = 1.5\beta$ (d) $\gamma = 1.95\beta$

Figure 8: Global and local convergence property of GFB method applied to the PCP problem. The setting of the problem is: the matrix size is 400×300, the $\text{rank}(X_{L,0}) = 10$, the sparsity of $X_{S,0}$ is 25% (25% of the elements of $X_{S,0}$ are non-zero), and the noise level is $\sigma = 0.01\text{std}(M)$.

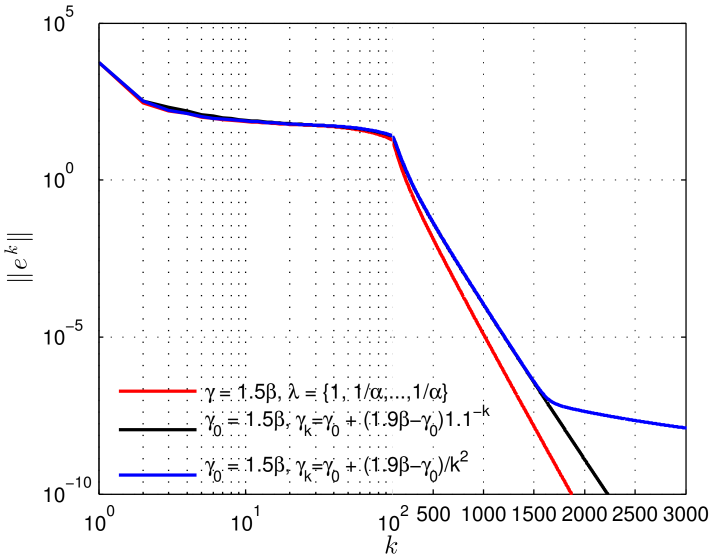


Figure 9: Comparison of stationary and non-stationary iteration of GFB.

8 Conclusion

In this paper, we present global iteration–complexity bounds for the inexact Krasnosel'skiĭ–Mann iteration built from non-expansive operator. Under metric subregularity, we also provided a unified quantitative analysis of local linear convergence. Extension to the non-stationary version is also proposed. The obtained results are applied to several monotone operator splitting algorithms and illustrated through several examples including matrix completion and PCP problems, where both global sublinear and local linear convergence profiles are observed. The local linear convergence rate depends on the subregularity modulus of the fixed point operator, which is not straightforward to compute in general. This is an important perspective that we will investigate in a future work.

Acknowledgements

This work has been partly supported by the European Research Council (ERC project SIGMA-Vision). We would like to thank Yuchao Tang for pointing [53] to us.

A Preparatory lemmas

Before we present the proofs of the main theorems in Section 3, we need some preparatory lemmas.

Lemma A.1. *For the error term e^k, the following inequality holds*

$$\frac{1}{2\lambda_k}\|e^k - e^{k+1}\|^2 \leq \langle e^k - \varepsilon^k, e^k - e^{k+1}\rangle.$$

Proof. Define $E : \mathcal{H} \to \mathcal{H}, z \mapsto (\mathrm{Id} - T)z$, as T is non-expansive, then $\frac{1}{2}E = \frac{1}{2}\big(\mathrm{Id} + (-T)\big) \in \mathcal{A}(\frac{1}{2})$ is firmly non-expansive (Lemma 2.2), therefore for $\forall p, q \in \mathcal{H}$, we have

$$\|\tfrac{1}{2}E(p) - \tfrac{1}{2}E(q)\|^2 \leq \langle p - q, \tfrac{1}{2}E(p) - \tfrac{1}{2}E(q)\rangle.$$

substituting z^k and z^{k+1} for p, q and combining the definition of e^k yield the result. ☐

Corollary A.2. *If T is α–averaged, then*

$$\frac{1}{2\alpha\lambda_k}\|e^k - e^{k+1}\|^2 \le \langle e^k - \varepsilon^k, e^k - e^{k+1}\rangle.$$

Lemma A.3. *For $z^\star \in \mathrm{Fix}(T)$, $\lambda_k \in]0,1]$, we have*

$$\|z^{k+1} - z^\star\|^2 \le \|z^k - z^\star\|^2 - \tau_k\|e^k\|^2 + \nu_1\lambda_k\|\varepsilon^k\|.$$

Proof. Proposition (3.1) and by virtue of [2, Corollary 2.14] we get

$$\begin{aligned}\|z^{k+1} - z^\star\|^2 &= \|T_{\lambda_k}z^k - z^\star + \lambda_k\varepsilon^k\|^2 \le \big(\|T_{\lambda_k}z^k - z^\star\| + \lambda_k\|\varepsilon^k\|\big)^2 \\ &\le \|(1-\lambda_k)(z^k - z^\star) + \lambda_k(Tz^k - Tz^\star)\|^2 + \nu_1\lambda_k\|\varepsilon^k\| \\ &= (1-\lambda_k)\|z^k - z^\star\|^2 + \lambda_k\|Tz^k - z^\star\|^2 - \tau_k\|z^k - Tz^k\|^2 + \nu_1\lambda_k\|\varepsilon^k\| \\ &\le \|z^k - z^\star\|^2 - \tau_k\|e^k\|^2 + \nu_1\lambda_k\|\varepsilon^k\|.\end{aligned}$$

□

This lemma indicates that sequence $\big(z^k\big)_{k\in\mathbb{N}}$ is quasi–Fejér monotone with respect to fix$T$.

Corollary A.4. *If T is α–averaged, then we have $\lambda_k \in]0, 1/\alpha]$ and*

$$\|z^{k+1} - z^\star\|^2 \le \|z^k - z^\star\|^2 - \lambda_k\left(\frac{1}{\alpha} - \lambda_k\right)\|e^k\|^2 + \nu_1\lambda_k\|\varepsilon^k\|\ .$$

Lemma A.5. *For $\lambda_k \in]0,1]$, sequence $(e^k)_{k\in\mathbb{N}}$ obeys*

$$\|e^{k+1}\|^2 - \nu_2\|\varepsilon^k\| \le \|e^k\|^2.$$

Proof. With the conclusion of Lemma A.1, we have

$$\begin{aligned}\|e^{k+1}\|^2 &= \|e^{k+1} - e^k + e^k\|^2 = \|e^k\|^2 - 2\langle e^k, e^k - e^{k+1}\rangle + \|e^k - e^{k+1}\|^2 \\ &\le \|e^k\|^2 - 2\langle e^k, e^k - e^{k+1}\rangle + 2\lambda_k\langle e^k - \varepsilon^k, e^k - e^{k+1}\rangle \\ &\le \|e^k\|^2 - 2(1-\lambda_k)\langle e^k - \varepsilon^k, e^k - e^{k+1}\rangle - 2\langle \varepsilon^k, e^k - e^{k+1}\rangle \\ &\le \|e^k\|^2 - \frac{1-\lambda_k}{\lambda_k}\|e^k - e^{k+1}\|^2 + 2\|e^k - e^{k+1}\|\|\varepsilon^k\| \\ &\le \|e^k\|^2 - \frac{1-\lambda_k}{\lambda_k}\|e^k - e^{k+1}\|^2 + \nu_2\|\varepsilon^k\| \le \|e^k\|^2 + \nu_2\|\varepsilon^k\|.\end{aligned}$$

□

The second last equation indicates that if the fixed point iteration (3.1) is exact, then sequence $\big(\|e^k\|^2\big)_{k\in\mathbb{N}}$ is monotone non-increasing, which is sharper than the result of [22, Proposition 11], also, if $\lambda_k = 1$, then $\big(\|z^k - z^{k+1}\|^2\big)_{k\in\mathbb{N}}$ is also monotone non-increasing. Moreover, if T is α–averaged, then $\lambda_k \in]0, 1/\alpha]$ and

$$\|e^{k+1}\|^2 \le \|e^k\|^2 - \frac{1-\alpha\lambda_k}{\alpha\lambda_k}\|e^k - e^{k+1}\|^2.$$

B Proofs of Section 3

Proof of Theorem 3.3. Condition (3.3) implies that $\underline{\tau} > 0$, $(\tau_k)_{k\in\mathbb{N}} \notin \ell_+^1$ and $\big(\lambda_k\|\varepsilon^k\|\big)_{k\in\mathbb{N}} \in \ell_+^1$, therefore, the iteration converges (Remark 3.2 (iii)). Moreover, we have $(\|e^k\|)_{k\in\mathbb{N}}$ and $(\|z^k - z^\star\|)_{k\in\mathbb{N}}$ are bounded, and ν_1 and ν_2 are bounded constants. Then from Lemma A.3, $\forall k \in \mathbb{N}$,

$$\tau_k\|e^k\|^2 \le \|z^k - z^\star\|^2 - \|z^{k+1} - z^\star\|^2 + \nu_1\lambda_k\|\varepsilon^k\|,$$

sum up from $j = 0$ to k,

$$\sum_{j=0}^{k} \tau_j\|e^j\|^2 \le \|z^0 - z^\star\|^2 - \|z^{k+1} - z^\star\|^2 + \nu_1 \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\|, \tag{B.1}$$

then from Lemma A.5, for $\forall j \le k$,

$$\|e^k\|^2 - \nu_2 \sum_{\ell=j}^{k-1} \|\varepsilon^\ell\| \le \|e^j\|^2.$$

substitute this into (B.1) we get,

$$\begin{aligned}\Big(\sum_{j=0}^{k} \tau_j\Big)\|e^k\|^2 &\le \sum_{j=0}^{k} \tau_j\|e^j\|^2 + \nu_2 \sum_{j=0}^{k} \tau_j \sum_{\ell=j}^{k-1} \|\varepsilon^\ell\| \\ &\le d_0^2 + \nu_1 \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\| + \nu_2 \sum_{j=0}^{k} \tau_j \sum_{\ell=j}^{k-1} \|\varepsilon^\ell\|,\end{aligned} \tag{B.2}$$

finally since $(k+1)\underline{\tau} \le \sum_{j=0}^{k} \tau_j$, then we have,

$$(k+1)\underline{\tau}\|e^k\|^2 \le d_0^2 + \nu_1 \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\| + \nu_2\overline{\tau} \sum_{\ell=0}^{k-1} (\ell+1)\|\varepsilon^\ell\|,$$

which leads to (3.4).

$\lambda_k \in [\frac{1}{2}, 1[$ is non-decreasing implies that τ_k is non-increasing, hence from (B.2) we get (3.5). □

Proof of Theorem 3.6. From (3.1), since T_{λ_k} is non-expansive, then

$$\begin{aligned}\|z^{k+1} - z^\star\| &= \|T_{\lambda_k} z^k - T_{\lambda_k} z^\star + \lambda_k\varepsilon^k\| \le \|z^k - z^\star\| + \lambda_k\|\varepsilon^k\| \\ &\le \|T_{\lambda_{k-1}} z^{k-1} - T_{\lambda_{k-1}} z^\star + \lambda_{k-1}\varepsilon^{k-1}\| + \lambda_k\|\varepsilon^k\| \\ &\le \|z^{k-1} - z^\star\| + \sum_{j=k-1}^{k} \lambda_j\|\varepsilon^j\| \le \|z^0 - z^\star\| + \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\|,\end{aligned}$$

together with the definition of $\bar{e}^k$, we have

$$\begin{aligned}\|\bar{e}^k\| &= \|\tfrac{1}{\Lambda_k} \textstyle\sum_{j=0}^{k} \lambda_j e^j\| = \frac{1}{\Lambda_k}\| \textstyle\sum_{j=0}^{k} (z^j - z^{j+1}) + \sum_{j=0}^{k}\lambda_j\varepsilon^j\| \\ &\le \frac{1}{\Lambda_k}\Big(\|z^0 - z^\star + z^\star - z^{k+1}\| + \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\|\Big) \\ &\le \frac{1}{\Lambda_k}\Big(\|z^0 - z^\star\| + \|z^\star - z^{k+1}\| + \sum_{j=0}^{k} \lambda_j\|\varepsilon^j\|\Big) \le \frac{2(d_0 + C_2)}{\Lambda_k}.\end{aligned}$$

□

Proof of Corollary 3.7. (i) By definition $v^k = \lambda_k e^k$, then from (3.4) we have

$$\|v^k\| = \|\lambda_k e^k\| \le \lambda_k \frac{d_0}{\sqrt{\underline{\tau}(k+1)}} \le \frac{d_0}{\sqrt{\underline{\tau}(k+1)}}.$$

(ii) This is a direct result of Theorem 3.6, replacing Λ_k with $k+1$ yields the desired result. □

C Proofs of Section 4

Proof of Theorem 4.2. Let constants $b > a > 0$ such that operator T' is metrically subregular at $z^\star$ with constant κ and ball $\mathbb{B}_a(z^\star)$. Make radius a smaller if necessary so that $\mathbb{B}_a(z^\star) \subset \mathbb{B}_b(z^\star) \subseteq \mathcal{Z}$ and

$$a + C_2 \le b.$$

Pick $z^0 \in \mathbb{B}_a(z^\star)$. If $z^0 = z^\star$ then, take $z^k = z^\star$ for all $k \in \mathbb{N}$ and there is nothing more to prove. If not, then from Lemma A.3 that for any $\tilde{z} \in \mathrm{fix}T$,

$$\|z^{k+1} - \tilde{z}\| \le \|z^k - \tilde{z}\| + \lambda_k \|\varepsilon^k\| \le \cdots \le \|z^0 - \tilde{z}\| + \sum_{j=0}^{k} \lambda_j \|\varepsilon^j\| \le a + C_2 \le b,$$

which implies that start from any point $z^0 \in \mathbb{B}_a(z^\star)$, $z^k \in \mathbb{B}_b(z^\star)$ holds for all $k \in \mathbb{N}$. Now for $\forall k \in \mathbb{N}$, let $\tilde{z} \in \mathrm{fix}T$ be such that $d_k = \|z^k - \tilde{z}\|$ and $c_k = \nu_1 \lambda_k \|\varepsilon^k\|$, then by virtue of the metric subregularity of T' and Lemma A.3, we have

$$\begin{aligned} d_{k+1}^2 &\le \|z^{k+1} - \tilde{z}\|^2 \le \|z^k - \tilde{z}\|^2 - \tau_k \|T' z^k - T' \tilde{z}\|^2 + c_k \\ &\le d_k^2 - \frac{\tau_k}{\kappa^2} d_k^2 + c_k \qquad \text{(C.1)} \\ &\le d_k^2 - \frac{\tau_k}{\kappa^2} \big(d_{k+1}^2 - c_k\big) + c_k \\ &= d_k^2 - \frac{\tau_k}{\kappa^2} d_{k+1}^2 + \frac{1}{\zeta_k} c_k. \qquad \text{(C.2)} \end{aligned}$$

If $\tau_k/\kappa^2 \in]0,1[$, then from (C.1) we have

$$d_{k+1}^2 \le \Big(1 - \frac{\tau_k}{\kappa^2}\Big) d_k^2 + c_k \ ,$$

or if $1 \le \tau_k/\kappa^2$, (C.2) produces

$$d_{k+1}^2 \le \frac{\kappa^2}{\kappa^2 + \tau_k} d_k^2 + c_k \ ,$$

$\lambda_k \in]0,1]$ ensures $\kappa^2/(\kappa^2 + \tau_k) \in]0,1]$. Therefore, we have

$$\zeta_k = \begin{cases} 1 - \frac{\tau_k}{\kappa^2}, & \text{if } \tau_k/\kappa^2 \in]0,1[\\ \frac{\kappa^2}{\kappa^2 + \tau_k}, & \text{if } 1 \le \tau_k/\kappa^2 \end{cases} \in]0,1].$$

Furthermore,

$$d_{k+1}^2 \le \zeta_k\, d_k^2 + c_k \le \cdots \le \chi_k\, d_0^2 + \sum_{j=0}^{k} \phi_{k-j} c_j \le \chi_k\, d_0^2 + \sum_{j=0}^{k} c_j, \qquad \text{(C.3)}$$

where $\chi_k = \prod_{j=0}^{k} \zeta_j$ and $\phi_{k-j} = \prod_{\ell=j+1}^{k} \zeta_\ell$.

(i) From (C.3) we have

$$d_{k+1}^2 \le d_k^2 + c_k,$$

then the $d_k^2 \to d \ge 0$ ([56, Chapter 2.2, Lemma 2]). If $(\tau_k)_{k\in\mathbb{N}} \notin \ell_+^1$, then $\|\boldsymbol{e}^k\| \to 0$ (Theorem 3.3), and by metric subregularity we have $d_k \le \kappa\|\boldsymbol{e}^k\|$, therefore $d = 0$;

(ii) If $\chi = \limsup_{k\to+\infty} \sqrt[k]{\chi_k} < 1$, then $\lim_{k\to+\infty} \chi_k = 0$ and sequences $(\chi_k)_{k\in\mathbb{N}}$, $(\phi_k)_{k\in\mathbb{N}} \in \ell_+^1$. Since also $(c_k)_{k\in\mathbb{N}} \in \ell_+^1$, therefore the convolution sequence $\big(\sum_{j=0}^k \phi_{k-j} c_j\big)_{k\in\mathbb{N}} \in \ell_+^1$, as a result, $\big(\chi_k\, d_0^2 + \sum_{j=0}^k \phi_{k-j} c_j\big)_{k\in\mathbb{N}} \in \ell_+^1$ and so is $(d_k^2)_{k\in\mathbb{N}}$.

If $\varepsilon^k = 0$, then from the inequality (C.3) we have $\lim_{k\to+\infty} \sqrt[k]{d_k} \le \limsup_{k\to+\infty} \sqrt[k]{\chi_k} < 1$.

(iii) If $0 < \inf_{k\in\mathbb{N}} \lambda_k \le \sup_{k\in\mathbb{N}} \lambda_k < 1$, then there exists $\zeta = \sup_{k\in\mathbb{N}} \kappa^2/(\kappa^2 + \tau_k) < 1$ which concludes the result.

(iv) If the set of fixed points $\mathrm{fix}T = \{z^\star\}$ is a singleton, then $d_k = \|z^k - z^\star\|$, and we can obtain the result from the above statements.

□

D Proofs of Section 5

D.1 Generalized Forward–Backward splitting

Proof of Proposition 5.2. Start from the Algorithm 1, we have

$$\begin{aligned}
\boldsymbol{z}^{k+1} &= \boldsymbol{z}^k + \lambda_k\Big[J_{\gamma\boldsymbol{A}}\big((2P_{\boldsymbol{\mathcal{S}}} - 2\gamma\boldsymbol{B}_{\boldsymbol{\mathcal{S}}} - \mathbf{Id} + \gamma\boldsymbol{B}_{\boldsymbol{\mathcal{S}}})\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k\big) \\
&\qquad - \frac{1}{2}\big((2P_{\boldsymbol{\mathcal{S}}} - 2\gamma\boldsymbol{B}_{\boldsymbol{\mathcal{S}}} - \mathbf{Id} + \gamma\boldsymbol{B}_{\boldsymbol{\mathcal{S}}})\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k\big) + \frac{1}{2}(\boldsymbol{z}^k - \gamma\boldsymbol{B}_{\boldsymbol{\mathcal{S}}}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k - \boldsymbol{z}^k\Big] \\
&= \boldsymbol{z}^k + \lambda_k\Big[J_{\gamma\boldsymbol{A}}R_{\boldsymbol{\mathcal{S}}}\big(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k\big) - \frac{1}{2}R_{\boldsymbol{\mathcal{S}}}\big(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k\big) + \frac{1}{2}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k - \boldsymbol{z}^k\Big] \\
&= \boldsymbol{z}^k + \lambda_k\Big[\frac{1}{2}(2J_{\gamma\boldsymbol{A}} - \mathbf{Id})R_{\boldsymbol{\mathcal{S}}}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \frac{1}{2}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k - \boldsymbol{z}^k\Big] \\
&= \boldsymbol{z}^k + \lambda_k\big(\boldsymbol{T}_{1,\gamma}(\boldsymbol{T}_{2,\gamma}\boldsymbol{z}^k + \boldsymbol{\varepsilon}_1^k) + \boldsymbol{\varepsilon}_2^k - \boldsymbol{z}^k\big).
\end{aligned}$$

□

Proof of Proposition 5.3. (i) For $\boldsymbol{v}^k$, the proof is the same as the Proof of Corollary 3.7 only that the bound for $\|\boldsymbol{e}^k\|$ is given by Corollary 3.5. Since $v = \sum_i \omega_i v_i$, then

$$\|v^k\| = \|\textstyle\sum_i \omega_i v_i\| \le \|\boldsymbol{v}^k\| \le \lambda_k\sqrt{\dfrac{d_0^2 + C_1}{\underline{\tau}(k+1)}},$$

since $(\lambda_k)_{k\in\mathbb{N}} \in [\frac{1}{2\alpha}, \frac{1}{\alpha}[$ is non-decreasing, then $\underline{\tau} = \tau_k = \lambda_k\big(\frac{1}{\alpha} - \lambda_k\big)$, which concludes the proof.

(ii) From Proposition 5.1 (ii), we have the following equivalence for equation (5.2),

$$
\begin{aligned}
&\boldsymbol{z}^{k+1} = \boldsymbol{z}^k + \lambda_k \Big(\big(\mathbf{Id} + \boldsymbol{A}'_\gamma\big)^{-1} (\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}}) \boldsymbol{z}^k + \boldsymbol{\varepsilon}^k - \boldsymbol{z}^k \Big) \\
\iff &\frac{\boldsymbol{z}^{k+1} - (1-\lambda_k)\boldsymbol{z}^k}{\lambda_k} - \boldsymbol{\varepsilon}^k = \big(\mathbf{Id} + \boldsymbol{A}'_\gamma\big)^{-1} (\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}}) \boldsymbol{z}^k \\
\iff &\frac{1}{\gamma}\boldsymbol{z}^k - \boldsymbol{B}_{\mathcal{S}}\boldsymbol{z}^k - \frac{1}{\gamma}\boldsymbol{p}^{k+1} \in \frac{1}{\gamma}\boldsymbol{A}'_\gamma \boldsymbol{p}^{k+1} \\
\iff &\frac{1}{\gamma}\boldsymbol{z}^k - \boldsymbol{B}_{\mathcal{S}}\boldsymbol{z}^k - \frac{1}{\gamma}\boldsymbol{p}^{k+1} + \boldsymbol{B}_{\mathcal{S}}\boldsymbol{p}^{k+1} \in \Big(\frac{1}{\gamma}\boldsymbol{A}'_\gamma + \boldsymbol{B}_{\mathcal{S}}\Big)\boldsymbol{p}^{k+1}.
\end{aligned}
$$

From Lemma 2.11 (ii), $\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}} \in \mathcal{A}(\frac{\gamma}{2\beta})$ is non-expansive, then

$$
\begin{aligned}
&d\Big(0, \frac{1}{\gamma}\boldsymbol{A}'_\gamma \boldsymbol{p}^{k+1} + \boldsymbol{B}_{\mathcal{S}}\boldsymbol{p}^{k+1}\Big) \\
&\le \|\boldsymbol{g}^{k+1} + \boldsymbol{B}_{\mathcal{S}}\boldsymbol{p}^{k+1}\| = \frac{1}{\gamma}\|(\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}})\boldsymbol{z}^k - (\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}})\boldsymbol{p}^{k+1}\| \\
&\le \frac{1}{\gamma}\|\boldsymbol{z}^k - \boldsymbol{p}^{k+1}\| = \frac{1}{\gamma}\|(\boldsymbol{z}^k - \boldsymbol{z}^{k+1})/\lambda_k + \boldsymbol{\varepsilon}^k\| = \frac{1}{\gamma}\|\boldsymbol{e}^k\| \le \frac{1}{\gamma}\sqrt{\frac{d_0^2 + C_1}{\tau_k(k+1)}}.
\end{aligned}
$$

□

Proof of Proposition 5.4. (i) For $\bar{\boldsymbol{v}}^k$, the proof is the same as the Proof of 3.7, as $\bar{v} = \sum_i \omega_i \bar{v}_i$, then

$$
\|\bar{v}^k\| = \| \textstyle\sum_i \omega_i \bar{v}_i^k\| \le \|\bar{\boldsymbol{v}}^k\| \le \dfrac{2(d_0 + C_2)}{k+1}.
$$

(ii) From Lemma 2.11 (ii), $\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}} \in \mathcal{A}(\frac{\gamma}{2\beta})$, and $\frac{1}{\lambda_0} \ge \frac{1}{\lambda_j}$, $j = 0, ..., k$ as $(\lambda_k)_{k\in\mathbb{N}}$ is non-decreasing, then from Theorem 3.6,

$$
\begin{aligned}
\|\bar{\boldsymbol{g}}^{k+1} + \boldsymbol{B}_{\mathcal{S}}\bar{\boldsymbol{p}}^{k+1}\| &= \frac{1}{\gamma}\|(\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}})\bar{\boldsymbol{z}}^k - (\mathbf{Id} - \gamma \boldsymbol{B}_{\mathcal{S}})\bar{\boldsymbol{p}}^{k+1}\| \\
&\le \frac{1}{\gamma}\|\bar{\boldsymbol{z}}^k - \bar{\boldsymbol{p}}^{k+1}\| = \frac{1}{\gamma(k+1)}\| \textstyle\sum_{j=0}^k \frac{1}{\lambda_j}(\boldsymbol{z}^j - \boldsymbol{z}^{j+1} + \lambda_j \boldsymbol{\varepsilon}^j)\| \\
&\le \frac{1}{\gamma\lambda_0(k+1)}\Big(\|\boldsymbol{z}^0 - \boldsymbol{z}^{k+1}\| + \sum_{j=0}^k \lambda_j \|\boldsymbol{\varepsilon}^j\|\Big) \le \frac{2(d_0 + C_2)}{\gamma\lambda_0(k+1)}.
\end{aligned}
$$

□

Proof of Proposition 5.5. Start from the definition of u_i^{k+1}, we have

$$
\begin{aligned}
&u_i^{k+1} = J_{\frac{\gamma}{\omega_i}A_i}(2x^k - z_i^k - \gamma B x^k) \\
\iff &2x^k - z_i^k - \gamma B x^k - u_i^{k+1} \in \frac{\gamma}{\omega_i} A_i u_i^{k+1} \\
\iff &\frac{\omega_i}{\gamma}\Big(2x^k - z_i^k - \gamma B x^k - u_i^{k+1} + \gamma B\big(\textstyle\sum_i \omega_i u_i^{k+1}\big)\Big) \in A_i u_i^{k+1} + \omega_i B\big(\textstyle\sum_i \omega_i u_i^{k+1}\big),
\end{aligned}
$$

then sum up over i,

$$
\frac{1}{\gamma}x^k - Bx^k - \frac{1}{\gamma}\textstyle\sum_i \omega_i u_i^{k+1} + B\big(\textstyle\sum_i \omega_i u_i^{k+1}\big) \in \textstyle\sum_i A_i u_i^{k+1} + B\big(\textstyle\sum_i \omega_i u_i^{k+1}\big),
$$

Since $\mathrm{Id} - \gamma B \in \mathcal{A}(\frac{\gamma}{2\beta})$ (Lemma 2.11 (ii)), then

$$\begin{aligned}
& d\Big(0, \textstyle\sum_i A_i u_i^{k+1} + B\big(\sum_i \omega_i u_i^{k+1}\big)\Big) \\
& \leq \|g^{k+1} + \gamma B\big(\textstyle\sum_i \omega_i u_i^{k+1}\big)\| = \frac{1}{\gamma}\|(\mathrm{Id} - \gamma B)x^k - (\mathrm{Id} - \gamma B)(\textstyle\sum_i \omega_i u_i^{k+1})\| \\
& \leq \frac{1}{\gamma}\|x^k - \textstyle\sum_i \omega_i u_i^{k+1}\| \leq \frac{1}{\gamma}\|\boldsymbol{x}^k - \boldsymbol{u}^{k+1}\| = \frac{1}{\gamma}\|\boldsymbol{e}^k\| \leq \frac{1}{\gamma}\sqrt{\frac{d_0^2 + C_1}{\tau_k(k+1)}}.
\end{aligned}$$

□

Proof of Proposition 5.6. Owing to Theorem 5.4, we have

$$\begin{aligned}
& \|\bar{g}^{k+1} + B(\textstyle\sum_i \omega_i \bar{u}_i^{k+1})\| \\
& = \frac{1}{\gamma}\|(\mathrm{Id} - \gamma B)\bar{x}^k - (\mathrm{Id} - \gamma B)(\textstyle\sum_i \omega_i \bar{u}_i^{k+1})\| \leq \frac{1}{\gamma}\|\bar{x}^k - \textstyle\sum_i \omega_i \bar{u}_i^{k+1}\| \\
& \leq \frac{1}{\gamma}\|\bar{\boldsymbol{x}}^k - \bar{\boldsymbol{u}}^{k+1}\| = \frac{1}{\gamma(k+1)}\|\textstyle\sum_{j=0}^k \big(\boldsymbol{z}^{j+1} - \boldsymbol{z}^j - \lambda_j \boldsymbol{\varepsilon}^j\big)/\lambda_j\| \\
& \leq \frac{1}{\gamma\lambda_0(k+1)}\Big(\|\boldsymbol{z}^0 - \boldsymbol{z}^{k+1}\| + \sum_{j=0}^k \lambda_j \|\boldsymbol{\varepsilon}^j\|\Big) \leq \frac{2(d_0 + C_2)}{\gamma\lambda_0(k+1)}.
\end{aligned}$$

□

D.2 Douglas–Rachford splitting and ADMM

Proof of Proposition 5.8. Note that at the convergence, there holds $\lim_{k\to+\infty} u^k = \lim_{k\to+\infty} v^k = x^\star$. From Algorithm 2, we have

$$\begin{aligned}
u^{k+1} = J_{\gamma A_1}(2x^k - z^k) &\iff 2x^k - z^k - u^{k+1} \in \gamma A_1 u^{k+1} \\
v^{k+1} = J_{\gamma A_2}(z^{k+1}) &\iff z^{k+1} - v^{k+1} \in \gamma A_2 v^{k+1}
\end{aligned}$$

Sum up the two inclusion, we get

$$\begin{aligned}
g^{k+1} &= \frac{1}{\gamma}\Big(2x^k - z^k - u^{k+1} + z^{k+1} - v^{k+1}\Big) \\
&= \frac{1}{\gamma}\Big(2x^k - z^k - \big(\tfrac{1}{\lambda_k}(z^{k+1} - z^k) + x^k - \varepsilon_1^k\big) + z^{k+1} - v^{k+1}\Big) \\
&= \frac{1}{\gamma}\Big((v^k - z^k) - (v^{k+1} - z^{k+1}) + \tfrac{1}{\lambda_k}(z^k - z^{k+1}) + \varepsilon_{1,2}^k\Big) \in A_1 u^{k+1} + A_2 x^{k+1},
\end{aligned}$$

where $\varepsilon_{1,2}^k = \varepsilon_1^k + \varepsilon_2^k$. Note that $v^k - z^k = \big(\mathrm{Id} - J_{\gamma A_2}\big)z^k$ and $\big(\mathrm{Id} - J_{\gamma A_2}\big)$ is firmly non-expansive (Lemma 2.2), therefore

$$\begin{aligned}
& d\Big(0, A_1 u^{k+1} + A_2 v^{k+1}\Big) \leq \|g^{k+1}\| \\
& = \frac{1}{\gamma}\|(v^k - z^k) - (v^{k+1} - z^{k+1}) + (z^k - z^{k+1})/\lambda_k + \varepsilon^k + \varepsilon_{1,2}^k - \varepsilon^k\| \\
& \leq \frac{1}{\gamma}\Big(\|z^k - z^{k+1}\| + \|e^k\| + \|\varepsilon_{1,2}^k - \varepsilon^k\|\Big) \leq \frac{1}{\gamma}\Big(\|z^k - z^{k+1} + \lambda_k\varepsilon^k - \lambda_k\varepsilon^k\| + \|e^k\|\Big) \\
& \leq \frac{1}{\gamma}\Big(\lambda_k\|e^{k+1}\| + \lambda_k\|\varepsilon^k\| + \|e^k\| + \|\varepsilon_{1,2}^k - \varepsilon^k\|\Big) \leq \frac{1+\lambda_k}{\gamma}\|e^k\| + \frac{1}{\gamma}\big(\lambda_k\|\varepsilon^k\| + \|\varepsilon_{1,2}^k - \varepsilon^k\|\big). \quad \text{(D.1)}
\end{aligned}$$

For the error $\|\varepsilon^k\|$, we have

$$\begin{aligned}\|\varepsilon^k\| &= \|\big(\tfrac{1}{2}\big(R_{\gamma A_1}(R_{\gamma A_2}\cdot + 2\varepsilon_2^k) + \mathrm{Id}\big)z^k - Tz^k\big) + \varepsilon_1^k\| \\ &\leq \frac{1}{2}\|R_{\gamma A_1}(R_{\gamma A_2}z^k + 2\varepsilon_2^k) - R_{\gamma A_1}R_{\gamma A_2}z^k\| + \|\varepsilon_1^k\| \\ &\leq \|\varepsilon_1^k\| + \|\varepsilon_2^k\|, \end{aligned} \tag{D.2}$$

and similiarly we have

$$\|\varepsilon^k - \varepsilon_{1,2}^k\| \leq 2\|\varepsilon_2^k\|. \tag{D.3}$$

Combining together (D.1), (D.2) and (D.3) concludes the proof. □

D.3 Vũ's primal–dual splitting

Proof of Proposition 5.10. First, we show that $\|\cdot\|_{\boldsymbol{F}}$ is lower–/upper–bounded by $\|\cdot\|_{\boldsymbol{\mathcal{K}}}$, define operator $\boldsymbol{V}$

$$\boldsymbol{V} : \mathcal{H} \to \boldsymbol{\mathcal{G}},\ x \mapsto \big(\sqrt{\sigma_1}^{-1}L_1x, \cdots, \sqrt{\sigma_n}^{-1}L_nx\big)$$

then, for $\forall x \in \mathcal{H}$,

$$\|\boldsymbol{V}x\|_{\boldsymbol{\mathcal{G}}}^2 = \sum_{i=1}^n \omega_i\sigma_i\|L_ix\|^2 \leq \|x\|^2\sum_{i=1}^n \omega_i\sigma_i\|L_i\|^2 \implies \|\boldsymbol{V}\|^2 \leq \sum_{i=1}^n \omega_i\sigma_i\|L_i\|^2,$$

set $\theta = \Big(\tau\sum_{i=1}^n \omega_i\sigma_i\|L_i\|^2\Big)^{-1/2} - 1$, then $\theta > 0$ and

$$\tau(1+\theta)\|\boldsymbol{V}\|^2 \leq \tau(1+\theta)\sum_{i=1}^n \omega_i\sigma_i\|L_i\|^2 = \frac{1}{1+\theta},$$

from the initial of Algorithm 3, we have $\tau\sum_i \omega_i\sigma_i\|L_i\|^2 < 1$, then for $\forall \boldsymbol{z} \in \boldsymbol{\mathcal{K}}$, we have

$$\begin{aligned}\langle \boldsymbol{z}, \boldsymbol{F}\boldsymbol{z}\rangle_{\boldsymbol{\mathcal{K}}} &= \tau^{-1}\|x\|^2 + \sum_{i=1}^n \sigma_i^{-1}\omega_i\|v_i\|^2 - 2\sum_{i=1}^n \omega_i\langle L_ix, v_i\rangle \\ &= \tau^{-1}\|x\|^2 + \sum_{i=1}^n \sigma_i^{-1}\omega_i\|v_i\|^2 - 2\sum_{i=1}^n \omega_i\langle \sqrt{\sigma_i}L_ix, \sqrt{\sigma_i}^{-1}v_i\rangle \\ &\leq \tau^{-1}\|x\|^2 + \sum_{i=1}^n \sigma_i^{-1}\omega_i\|v_i\|^2 + \sum_{i=1}^n \omega_i\Big(\sigma_i\|L_ix\|^2 + \sigma_i^{-1}\|v_i\|^2\Big) \\ &\leq \tau^{-1}\|x\|^2 + \tau^{-1}\|x\|^2\Big(\tau\sum_{i=1}^n \omega_i\sigma_i\|L_i\|^2\Big) + 2\sum_{i=1}^n \sigma_i^{-1}\omega_i\|v_i\|^2 \\ &\leq 2\Big(\tau^{-1}\|x\|^2 + \sum_{i=1}^n \sigma_i^{-1}\omega_i\|v_i\|^2\Big) \leq 2\delta\Big(\|x\|^2 + \sum_{i=1}^n \omega_i\|v_i\|^2\Big) \\ &= 2\delta\|\boldsymbol{z}\|_{\boldsymbol{\mathcal{K}}}, \end{aligned}$$

where $\delta = \max\big\{\tau^{-1}, {\sigma_1}^{-1}, ..., {\sigma_n}^{-1}\big\}$, and from [68, Equation 3.20], we have

$$\eta\|\boldsymbol{z}\|_{\boldsymbol{\mathcal{K}}} \leq \langle \boldsymbol{z}, \boldsymbol{F}\boldsymbol{z}\rangle_{\boldsymbol{\mathcal{K}}},$$

therefore,

$$\eta\|\boldsymbol{z}\|_{\boldsymbol{\mathcal{K}}} \leq \|\boldsymbol{z}\|_{\boldsymbol{F}} \leq 2\delta\|\boldsymbol{z}\|_{\boldsymbol{\mathcal{K}}}. \tag{D.4}$$

Define the distance $\tilde{d}_0 = \inf_{z^\star \in \mathrm{fix} T} \|z^0 - z^\star\|_F$, $\tilde{\nu}_1 = 2\sup_{k\in\mathbb{N}} \|T_k z^k - z^\star\|_F + \sup_{k\in\mathbb{N}} \lambda_k \|\varepsilon^k\|_F$, $\tilde{\nu}_2 = 2\sup_{k\in\mathbb{N}} \|e^k - e^{k+1}\|_F$, $\tilde{C}_1 = \nu_1 \sum_{j\in\mathbb{N}} \lambda_j \|\varepsilon^j\|_F + \nu_2 \overline{\tau} \sum_{\ell\in\mathbb{N}} (\ell+1)\|\varepsilon^\ell\|_F < +\infty$. From the bounds (D.4), we have

$$\tilde{d}_0 \le 2\delta d_0,\ \tilde{\nu}_1 \le 2\delta\nu_1,\ \text{and}\ \tilde{\nu}_2 \le 2\delta\nu_2. \tag{D.5}$$

For the fixed point iteration (5.8), under metric F, we have

$$\|e^k\|_F \le \sqrt{\frac{\tilde{d}_0^2 + \tilde{C}_1}{\underline{\tau}(k+1)}},\ \|v^k\|_F \le \lambda_k \sqrt{\frac{\tilde{d}_0^2 + \tilde{C}_1}{\underline{\tau}(k+1)}},\ \text{and}\ \|g_F^{k+1} + Bw^{k+1}\|_F \le \sqrt{\frac{\tilde{d}_0^2 + \tilde{C}_1}{\underline{\tau}(k+1)}},$$

then combine (D.4) and (D.5) we obtain the desired result. □

Proof of Proposition 5.12. The result of a combination of (D.4), Theorem 3.6, Corollary 3.7 and 5.4.

(i) For $\bar{e}^k$, we have

$$\begin{aligned}
\|\bar{e}^k\|_{\mathcal{K}} &\le \frac{1}{\eta}\|\bar{e}^k\|_F = \frac{1}{\eta}\|\tfrac{1}{\Lambda_k}\textstyle\sum_{j=0}^k \lambda_j e^j\|_F = \frac{1}{\eta\Lambda_k}\|\textstyle\sum_{j=0}^k (e^j - e^{j+1}) + \sum_{j=0}^k \lambda_j \varepsilon^j\|_F \\
&\le \frac{1}{\eta\Lambda_k}\Big(\|z^0 - z^\star + z^\star - z^{k+1}\|_F + \sum_{j=0}^k \lambda_j\|\varepsilon^j\|_F\Big) \\
&\le \frac{2\delta}{\eta\Lambda_k}\Big(\|z^0 - z^\star\|_{\mathcal{K}} + \|z^\star - z^{k+1}\|_{\mathcal{K}} + \sum_{j=0}^k \lambda_j\|\varepsilon^j\|_{\mathcal{K}}\Big) \le \frac{4\delta(d_0 + C_2)}{\eta\Lambda_k}.
\end{aligned}$$

Replacing Λ_k with $k+1$ obtains the result for $\|\bar{v}^k\|_{\mathcal{K}}$.

(ii) $\mathbf{Id} - \gamma B_{\mathcal{S}}$ is non-expansive (Lemma 2.11 (ii)), then from Theorem 3.6,

$$\begin{aligned}
&\|\bar{g}_F^{k+1} + B\bar{w}^{k+1}\|_{\mathcal{K}} \\
&\le \frac{1}{\eta}\|\bar{g}_F^{k+1} + B\bar{w}^{k+1}\|_F = \frac{1}{\eta\gamma}\|(\mathbf{Id} - \gamma B)\bar{z}^k - (\mathbf{Id} - \gamma B)\bar{w}^{k+1}\|_F \\
&\le \frac{1}{\eta\gamma}\|\bar{z}^k - \bar{w}^{k+1}\|_F = \frac{1}{\eta\gamma(k+1)}\|\textstyle\sum_{j=0}^k \frac{1}{\lambda_j}(z^j - z^{j+1} + \lambda_j\varepsilon^j)\|_F \\
&\le \frac{1}{\eta\gamma\underline{\lambda}(k+1)}\Big(\|z^0 - z^{k+1}\|_F + \sum_{j=0}^k \lambda_j\|\varepsilon^j\|_F\Big) \le \frac{4\delta(d_0 + C_2)}{\eta\gamma\underline{\lambda}(k+1)}.
\end{aligned}$$

□

Proof of Proposition 5.13. From Lemma 2.11 (ii) we have $\mathrm{Id} - \tau B \in \mathcal{A}(\frac{\tau}{2\mu})$ and $\mathrm{Id} - \sigma_i D_i^{-1} \in \mathcal{A}(\frac{\sigma_i}{2\nu_i})$, $i \in [\![1, n]\!]$ are non-expansive, denote them G_B and G_{D_i} respectively, then we have

$$\begin{aligned}
& d\Big(0, M\big(v^k, t^{k+1}, y^{k+1}\big)\Big) \le \|g^{k+1}\|_{\mathcal{K}} \\
&= \Big(\|\tfrac{1}{\tau}(G_B x^k - G_B s^{k+1})\|^2 + \sum_{i=1}^n \omega_i\|\tfrac{1}{\sigma_i}(G_{D_i} v_i^k - G_{D_i} t_i^{k+1})\|^2\Big)^{1/2} \\
&\le \Big(\frac{1}{\tau^2}\|x^k - s^{k+1}\|^2 + \sum_{i=1}^n \omega_i \frac{1}{\sigma_i^2}\|v_i^k - t_i^{k+1}\|^2\Big)^{1/2} \le \delta\Big(\|x^k - s^{k+1}\|^2 + \sum_{i=1}^n \omega_i\|v_i^k - t_i^{k+1}\|^2\Big)^{1/2} \\
&= \delta\|z^k - t^{k+1}\|_{\mathcal{K}} = \delta\|e^k\|_{\mathcal{K}} \le \frac{2\delta^2}{\eta}\sqrt{\frac{d_0^2 + C_1}{\underline{\tau}(k+1)}}.
\end{aligned}$$

□

E Proofs of Section 6

Proof of Theorem 6.1. Let $(z^k)_{k\in\mathbb{N}}$ be a sequence provided by the perturbed approximate method, and $z^\star \in \mathrm{fix}T_\Gamma$. Then

$$\begin{aligned}\|z^{k+1} - z^\star\| &= \|T_{\Gamma_k,\lambda_k} z^k + \varepsilon^k - T_{\Gamma,\lambda_k} z^\star\| \\ &\le \|T_{\Gamma_k,\lambda_k} z^k - T_{\Gamma_k,\lambda_k} z^\star\| + \|T_{\Gamma_k,\lambda_k} z^\star - T_{\Gamma,\lambda_k} z^\star\| + \|\varepsilon^k\| \\ &\le (1+\beta_k)\|z^k - z^\star\| + \Delta_{k,\|z^\star\|} + \|\varepsilon^k\|.\end{aligned}$$

As $(\varepsilon^k)_{k\in\mathbb{N}}$, $(\delta^k)_{k\in\mathbb{N}}$ and $(\Delta_{k,\|z^\star\|})_{k\in\mathbb{N}}$ are summable by assumptions (b), (d) and (e) of Theorem 6.1, it follows from [56, Lemma 2.2.2] that the sequence $(\|z^k - z^\star\|)_{k\in\mathbb{N}}$ converges, hence bounded. Therefore, z^k is bounded in norm by some $\rho \in [0,+\infty[$. This implies that

$$\|z^{k+1} - T_{\Gamma,\lambda_k} z^k\| = \|T_{\Gamma_k,\lambda_k} z^k + \varepsilon^k - T_{\Gamma,\lambda_k} z^k\| \le \Delta_{k,\rho} + \|\varepsilon^k\|,$$

the perturbed approximate method can be seen as an approximate version of the exact method with an error term which is summable owing to (e) of Theorem 6.1. The rest of the proof follows by applying [39, Lemma 2.1 and Remark 2.2] (see also [6, Remark 14]) using ((a)), (c) of Theorem 6.1. □

Proof of Theorem 6.3. Since

$$\sum_{k\in\mathbb{N}} (k+1)\|\pi^k\| < +\infty,$$

then following the proofs of Theorem 3.3 and 3.6 obtains the desired result. □

Proof of Theorem 6.4. As the assumptions required by [57, Theorem 4.1] for weak convergence of the exact method are fulfilled, it is sufficient to verify the conditions of Theorem 6.1 to conclude.

1. With the exact and perturbed operators, assumption (a) of Theorem 6.1 is verified as $\emptyset \ne \mathrm{zer}\big(B + \sum_i A_i\big) = P_{\boldsymbol{\mathcal{S}}}\big(\mathrm{fix}\boldsymbol{T}_\gamma\big)$;
2. as $\boldsymbol{T}_{\gamma_k,\lambda_k}$ is α_k–averaged, hence assumption 6.1 (b) holds;
3. assumption 6.1 (c) holds by assumption on λ_k;
4. assumption 6.1 (d) holds by assumption on $\boldsymbol{\varepsilon}^k$. And it is sufficient to check that the summability condition of 6.1 (e) is in force if 6.4 (d) holds and $(|\gamma_k - \gamma|)_{k\in\mathbb{N}}$ is summable;

For $\forall\rho \ge 0$, $\forall \boldsymbol{z}$ such that $\|\boldsymbol{z}\| \le \rho$, then we have

$$\begin{aligned}\|\boldsymbol{T}_{\gamma_k}\boldsymbol{z} - \boldsymbol{T}_\gamma\boldsymbol{z}\| &\le \big\|\big(\alpha_k\boldsymbol{R}_{\gamma_k} + (1-\alpha_k)\mathbf{Id}\big)\boldsymbol{z} - \big(\alpha\boldsymbol{R}_\gamma + (1-\alpha)\mathbf{Id}\big)\boldsymbol{z}\big\| \\ &\le \|\alpha_k\boldsymbol{z} - \alpha\boldsymbol{z}\| + \|\alpha_k\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \alpha\boldsymbol{R}_\gamma\boldsymbol{z}\| \\ &\le |\alpha_k - \alpha|\rho + \|\alpha_k\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \alpha\boldsymbol{R}_\gamma\boldsymbol{z}\|.\end{aligned} \tag{E.1}$$

The term $\|\alpha_k\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \alpha\boldsymbol{R}_\gamma\boldsymbol{z}\|$ needs to be bounded, and we have

$$\begin{aligned}\|\alpha_k\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \alpha\boldsymbol{R}_\gamma\boldsymbol{z}\| &= \|\alpha_k\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \alpha_k\boldsymbol{R}_\gamma\boldsymbol{z} + \alpha_k\boldsymbol{R}_\gamma\boldsymbol{z} - \alpha\boldsymbol{R}_\gamma\boldsymbol{z}\| \\ &= \alpha_k\|\boldsymbol{R}_{\gamma_k}\boldsymbol{z} - \boldsymbol{R}_\gamma\boldsymbol{z}\| + |\alpha_k - \alpha|\|\boldsymbol{R}_\gamma\boldsymbol{z} - \boldsymbol{R}_\gamma\mathbf{0} + \mathbf{R}_\gamma\mathbf{0}\| \\ &\le \alpha_k\Delta_{k,\rho} + |\alpha_k - \alpha|\big(\rho + \|\boldsymbol{R}_\gamma\mathbf{0}\|\big),\end{aligned} \tag{E.2}$$

where $\|\boldsymbol{R}_\gamma \mathbf{0}\|$ is bounded, and for $|\alpha_k - \alpha|$, we have

$$|\alpha_k - \alpha| = \Big|\frac{2\beta}{4\beta - \gamma_k} - \frac{2\beta}{4\beta - \gamma}\Big| = \Big|\frac{2\beta(\gamma - \gamma_k)}{(4\beta - \gamma_k)(4\beta - \gamma)}\Big| \leq \frac{1}{2\beta}|\gamma_k - \gamma|.$$

Therefore we get $\forall \rho \in [0, +\infty[$,

$$\sum_{n\in\mathbb{N}} \sup_{\|\boldsymbol{z}\|\leq\rho} \|\boldsymbol{T}_{\gamma_k}\boldsymbol{z} - \boldsymbol{T}_\gamma \boldsymbol{z}\| \leq \sum_{k\in\mathbb{N}} \alpha_k \Delta_{k,\rho} + C\sum_{k\in\mathbb{N}} |\gamma_k - \gamma| < +\infty$$

where $C = \frac{1}{2\beta}\big(2\rho + \|\boldsymbol{R}_\gamma \mathbf{0}\|\big) < +\infty$. Consequently, (e) of Theorem 6.1 is fulfilled.

When $\sum_{k\in\mathbb{N}}(k+1)\|\boldsymbol{\pi}^k\| < +\infty$, like Theorem 6.3, the problem boils down to the scenarios considered in Theorem 3.3 and 3.6, and we can easily obtain the corresponding iteration–complexity bounds. □

References

[1] J. B. Baillon and G. Haddad. Quelques propriétés des opérateurs angle-bornés etn-cycliquement monotones. *Israel Journal of Mathematics*, 26(2):137–150, 1977.

[2] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011.

[3] H. H. Bauschke, D. R. Luke, H. M. Phan, and X. Wang. Restricted normal cones and the method of alternating projections. arXiv 1205.0318, 2012.

[4] J. M. Borwein and B. Sims. The douglas-rachford algorithm in the absence of convexity. 2010.

[5] K. Bredies and D. A. Lorenz. Linear convergence of iterative soft-thresholding. *Journal of Fourier Analysis and Applications*, 14(5-6):813–837, 2008.

[6] H. Brézis and P. L. Lions. Produits infinis de résolvantes. *Israel Journal of Mathematics*, 29(4):329–345, 1978.

[7] L. M. Briceño-Arias. Forward-douglas-rachford splitting and forward-partial inverse method for solving monotone inclusions. *arXiv preprint arXiv:1212.5942*, 2012.

[8] L. M. Briceno-Arias and P. L. Combettes. A monotone+ skew splitting model for composite monotone inclusions in duality. *SIAM Journal on Optimization*, 21(4):1230–1250, 2011.

[9] R. S. Burachik, S. Scheimberg, and B. F. Svaiter. Robustness of the hybrid extragradient proximal-point algorithm. *Journal of Optimization Theory and Applications*, 111(1):117–136, 2001.

[10] E. J. Candès, X. Li, Y. Ma, and J. Wright. Robust principal component analysis? *Journal of the ACM (JACM)*, 58(3):11, 2011.

[11] E. J. Candès and B. Recht. Exact matrix completion via convex optimization. *Foundations of Computational Mathematics*, May 2009.

[12] E. J. Candès and T. Tao. The power of convex relaxation: near-optimal matrix completion. *IEEE Trans. Inform. Theory*, 56(5):2053–2080, 2010.

[13] A. Chambolle and J. Darbon. On total variation minimization and surface evolution using parametric maximum flows. *International journal of computer vision*, 84(3):288–307, 2009.

[14] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

[15] G. Chen and M. Teboulle. A proximal-based decomposition method for convex minimization problems. *Mathematical Programming*, 64(1-3):81–101, 1994.

[16] G. HG Chen and R. T. Rockafellar. Convergence rates in forward–backward splitting. *SIAM Journal on Optimization*, 7(2):421–444, 1997.

[17] P. Chen, J. Huang, and X. Zhang. A primal–dual fixed point algorithm for convex separable minimization with applications to image restoration. *Inverse Problems*, 29(2):025011, 2013.

[18] P. L. Combettes. Quasi-fejérian analysis of some optimization algorithms. *Studies in Computational Mathematics*, 8:115–152, 2001.

[19] P. L. Combettes. Solving monotone inclusions via compositions of nonexpansive averaged operators. *Optimization*, 53(5-6):475–504, 2004.

[20] 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 and variational analysis*, 20(2):307–330, 2012.

[21] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *Multiscale Modeling & Simulation*, 4(4):1168–1200, 2005.

[22] R. Cominetti, J. A. Soto, and J. Vaisman. On the rate of convergence of krasnoselski-mann iterations and their connection with sums of bernoullis. *Israel Journal of Mathematics*, pages 1–16, 2012.

[23] L. Condat. A primal–dual splitting method for convex optimization involving lipschitzian, proximable and linear composite terms. *Journal of Optimization Theory and Applications*, pages 1–20, 2012.

[24] L. Demanet and X. Zhang. Eventual linear convergence of the douglas-rachford iteration for basis pursuit. *Mathematical Programming*, 2013.

[25] A. L. Dontchev, M. Quincampoix, and N. Zlateva. Aubin criterion for metric regularity. *Journal of Convex Analysis*, 13:281–297, 2006.

[26] A. L. Dontchev and R. T. Rockafellar. *Implicit functions and solution mappings: A view from variational analysis.* Springer, 2009.

[27] J. Douglas and H. H. Rachford. On the numerical solution of heat conduction problems in two and three space variables. *Transactions of the American mathematical Society*, 82(2):421–439, 1956.

[28] J. Eckstein and D. P. Bertsekas. On the douglasñrachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.

[29] M. Fortin and R. Glowinski. *Augmented Lagrangian methods: applications to the numerical solution of boundary-value problems.* Access Online via Elsevier, 2000.

[30] D. Gabay. Chapter ix applications of the method of multipliers to variational inequalities. *Studies in mathematics and its applications*, 15:299–331, 1983.

[31] D. Gabay and B. Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications*, 2(1):17–40, 1976.

[32] R. Glowinski and P. Le Tallec. *Augmented Lagrangian and operator-splitting methods in nonlinear mechanics*, volume 9. SIAM, 1989.

[33] B. He and X. Yuan. On convergence rate of the douglas-rachford operator splitting method. *Mathematical Programming, under revision*, 2011.

[34] B. He and X. Yuan. On non-ergodic convergence rate of douglas-rachford alternating direction method of multipliers. Technical report, Technical report, Tech. rep, 2012.

[35] R. Hesse and D. R. Luke. Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. *arXiv preprint arXiv:1212.3349*, 2012.

[36] R. Hesse and D. R. Luke. Nonmonotone fixed point mappings and convergence of first-order algorithms. *Preprint*, 2012.

[37] D. S. Hochbaum. An efficient algorithm for image segmentation, markov random fields and related problems. *Journal of the ACM (JACM)*, 48(4):686–701, 2001.

[38] M. A. Krasnosel'skii. Two remarks on the method of successive approximations. *Uspekhi Matematicheskikh Nauk*, 10(1):123–127, 1955.

[39] B. Lemaire. Stability of the iteration method for non expansive mappings. *Serdica Mathematical Journal*, 22(3):331p–340p, 1996.

[40] B. Lemaire. Which fixed point does the iteration method select? In *Recent Advances in Optimization*, pages 154–167. Springer, 1997.

[41] D. Leventhal. Metric subregularity and the proximal point method. *Journal of Mathematical Analysis and Applications*, 360(2):681–688, 2009.

[42] A. Lewis, D. Luke, and J. Malick. Local linear convergence for alternating and averaged nonconvex projections. *Found. Comput. Math.*, 9(4):485–513, 2009.

[43] G. Li and B. S. Mordukhovich. Hölder metric subregularity with applications to proximal point method. *SIAM Journal on Optimization*, 22(4):1655–1684, 2012.

[44] P. L. Lions and B. Mercier. Splitting algorithms for the sum of two nonlinear operators. *SIAM Journal on Numerical Analysis*, 16(6):964–979, 1979.

[45] W. R. Mann. Mean value methods in iteration. *Proceedings of the American Mathematical Society*, 4(3):506–510, 1953.

[46] B. Martinet. Brève communication. régularisation d'inéquations variationnelles par approximations successives. *ESAIM: Mathematical Modelling and Numerical Analysis-Modélisation Mathématique et Analyse Numérique*, 4(R3):154–158, 1970.

[47] B. Mercier. Topics in finite element solution of elliptic problems. *Lectures on Mathematics*, 63, 1979.

[48] G. J. Minty. Monotone (nonlinear) operators in hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

[49] R. DC Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010.

[50] J. J. Moreau. Décomposition orthogonale dõun espace hilbertien selon deux cônes mutuellement polaires. *CR Acad. Sci. Paris*, 255:238–240, 1962.

[51] J. J. Moreau. Proximité et dualité dans un espace hilbertien. *Bulletin de la Société mathématique de France*, 93:273–299, 1965.

[52] Y. Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer, 2004.

[53] N. Ogura and I. Yamada. Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping. *Numerical Functional Analysis and Optimization*, 23(1-2):113–137, 2002.

[54] G. B. Passty. Ergodic convergence to a zero of the sum of monotone operators in hilbert space. *Journal of Mathematical Analysis and Applications*, 72(2):383–390, 1979.

[55] D. W. Peaceman and H. H. Rachford, Jr. The numerical solution of parabolic and elliptic differential equations. *Journal of the Society for Industrial & Applied Mathematics*, 3(1):28–41, 1955.

[56] B. T. Polyak. *Introduction to optimization.* Optimization Software, 1987.

[57] H. Raguet, J. Fadili, and G. Peyré. Generalized forward-backward splitting. *SIAM Journal on Imaging Sciences*, 6(3):1199–1226, 2013.

[58] B. Recht, M. Fazel, and P. A. Parrilo. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Review*, 52(3):471–501, 2010.

[59] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.

[60] R. T. Rockafellar. *Convex analysis*, volume 28. Princeton university press, 1997.

[61] L. I Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[62] M. V. Solodov. A class of decomposition methods for convex optimization and monotone variational inclusions via the hybrid inexact proximal point framework. *Optimization Methods and Software*, 19(5):557–575, 2004.

[63] M. V. Solodov and B. F. Svaiter. A hybrid approximate extragradient–proximal point algorithm using the enlargement of a maximal monotone operator. *Set-Valued Analysis*, 7(4):323–345, 1999.

[64] B. F. Svaiter. A class of fejer convergent algorithms, approximate resolvents and the hybrid proximal-extragradient method. *arXiv preprint arXiv:1204.1353*, 2012.

[65] P. Tseng. Applications of a splitting algorithm to decomposition in convex programming and variational inequalities. *SIAM Journal on Control and Optimization*, 29(1):119–138, 1991.

[66] P. Tseng. Alternating projection-proximal methods for convex programming and variational inequalities. *SIAM Journal on Optimization*, 7(4):951–965, 1997.

[67] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.

[68] B. C. Vũ. A splitting algorithm for dual monotone inclusions involving cocoercive operators. *Advances in Computational Mathematics*, pages 1–15, 2011.