CrossMark

A convergent relaxation of the Douglas–Rachford algorithm

Nguyen Hieu Thao[1,2] 🆔

Abstract This paper proposes an algorithm for solving structured optimization problems, which covers both the backward–backward and the Douglas–Rachford algorithms as special cases, and analyzes its convergence. The set of fixed points of the corresponding operator is characterized in several cases. Convergence criteria of the algorithm in terms of general fixed point iterations are established. When applied to nonconvex feasibility including potentially inconsistent problems, we prove local linear convergence results under mild assumptions on regularity of individual sets and of the collection of sets. In this special case, we refine known linear convergence criteria for the Douglas–Rachford (DR) algorithm. As a consequence, for feasibility problem with one of the sets being affine, we establish criteria for linear and sublinear convergence of convex combinations of the alternating projection and the DR methods. These results seem to be new. We also demonstrate the seemingly improved numerical performance of this algorithm compared to the RAAR algorithm for both consistent and inconsistent sparse feasibility problems.

✉ Nguyen Hieu Thao
  h.t.nguyen-3@tudelft.nl; hieuthao.ctu@gmail.com

[1] Delft Center for Systems and Control, Delft University of Technology, 2628CD Delft, The Netherlands

[2] Department of Mathematics, School of Education, Can Tho University, Can Tho, Vietnam

🍃 Springer

1 Introduction

Convergence analysis has been one of the central and very active applications of variational analysis and mathematical optimization. Examples of recent contributions to the theory of the field that have initiated efficient programs of analysis are [1,2,7,38]. It is the common recipe emphasized in these and many other works that there are two key ingredients required in order to derive convergence of a numerical method (1) regularity of individual functions or sets such as *convexity* and *averaging property*, and (2) regularity of collections of functions or sets at their critical points such as *transversality*, *Kurdyka-Łojasiewicz property* and *metric subregularity*. As a result, the question about convergence of a solving method can often be reduced to checking whether certain regularity properties of the problem data are satisfied. There have been a considerable number of papers studying these two ingredients of convergence analysis in order to establish sharper convergence criteria in various circumstances, especially those applicable to algorithms for solving nonconvex problems [5,12,13, 19,26,27,31–33,38,42,45].

This paper suggests an algorithm called T_λ, which covers both the backward-backward and the DR algorithms as special cases of choosing the parameter $\lambda \in [0, 1]$, and analyzes its convergence. When applied to feasibility problem for two sets one of which is affine, T_λ is a convex combination of the alternating projection and the DR methods. On the other hand, T_λ can be viewed as a relaxation of the DR algorithm. Motivation for relaxing the DR algorithm comes from the lack of stability of this algorithm when applied to inconsistent problems. This phenomenon has been observed for the *Fourier phase retrieval problem* which is essentially inconsistent due to the reciprocal relationship between the spatial and frequency variables of the Fourier transform [35,36]. To address this issue, a relaxation of the DR algorithm, often known as the RAAR algorithm, was proposed and applied to phase retrieval problems by Luke in the aforementioned papers. In the framework of feasibility, the RAAR algorithm is described as a convex combination of the basic DR operator and one of the projectors. Our preliminary numerical experiments have revealed a promising performance of algorithm T_λ in comparison with the RAAR method. This observation has motivated the study of convergence analysis of algorithm T_λ in this paper.

After introducing the notation and proving preliminary results in Sect. 2, we introduce T_λ as a general fixed point operator, characterize the set of fixed points of T_λ (Proposition 1), and establish abstract convergence criteria for iterations generated by T_λ (Theorem 2) in Sect. 3. We discuss algorithm T_λ in the framework of feasibility problems in Sect. 4. The set of fixed points of T_λ is characterized for convex inconsistent feasibility (Proposition 3). For consistent feasibility we show that almost

averagedness of T_λ (Proposition 4) and metric subregularity of $T_\lambda -$ Id (Lemma 3) can be obtained from regular properties of the individual sets and of the collection of sets, respectively. As a result, the two regularity notions are combined to yield local linear convergence of iterations generated by T_λ (Theorem 4). Section 5 is devoted to demonstrate the improved numerical performance of algorithm T_λ compared to the RAAR algorithm for both consistent and inconsistent feasibility problems. In this section, we study the feasibility approach for solving the *sparse optimization problem*. Our linear convergence result established in Sect. 4 for iterations generated by T_λ is also illustrated in this application (Theorem 5).

2 Notation and preliminary results

Our notation is standard, c.f. [11,40,46]. The setting throughout this paper is a finite dimensional Euclidean space $\mathbb{E}$. The norm $\|\cdot\|$ denotes the Euclidean norm. The open unit ball in a Euclidean space is denoted $\mathbb{B}$, and $\mathbb{B}_\delta(x)$ stands for the open ball with radius $\delta > 0$ and center x. The distance to a set $A \subset \mathbb{E}$ with respect to the bivariate function dist $(\cdot, \cdot)$ is defined by

$$\text{dist}\,(\cdot, A)\colon \mathbb{E} \to \mathbb{R}_+ \colon x \mapsto \inf_{y \in A} \text{dist}\,(x, y).$$

We use the convention that the distance to the empty set is $+\infty$. The set-valued mapping

$$P_A : \mathbb{E} \rightrightarrows \mathbb{E} : x \mapsto \{y \in A \mid \text{dist}\,(x, y) = \text{dist}\,(x, A)\}$$

is the *projector* on A. An element $y \in P_A(x)$ is called a *projection*. This exists for any closed set $A \subset \mathbb{E}$. Note that the projector is not, in general, single-valued. Closely related to the projector is the *prox* mapping corresponding to a function f and a stepsize $\tau > 0$ [41]

$$\text{prox}_{\tau, f}(x) := \text{argmin}_{y \in \mathbb{E}} \left\{ f(y) + \tfrac{1}{2\tau} \|y - x\|^2 \right\}.$$

When $f = \iota_A$ is the *indicator function* of A, that is $\iota_A(x) = 0$ if $x \in A$ and $\iota_A(x) = +\infty$ otherwise, then $\text{prox}_{\tau, \iota_A} = P_A$ for all $\tau > 0$. The *inverse of the projector*, P_A^{-1}, is defined by

$$P_A^{-1}(a) := \{x \in \mathbb{E} \mid a \in P_A(x)\}.$$

The *proximal normal cone* to A at $\bar{x}$ is the set, which need not be either closed or convex,

$$N_A^{\text{prox}}(\bar{x}) := \text{cone}\left(P_A^{-1}(\bar{x}) - \bar{x}\right). \tag{1}$$

If $\bar{x} \notin A$, then $N_A^{\text{prox}}(\bar{x})$ is defined to be empty. Normal cones are central to characterizations both of the regularity of individual sets and of the regularity of collections of sets. For a refined numerical analysis of projection methods, one also defines the Λ-*proximal normal cone* to A at $\bar{x}$ by

$$N_{A|\Lambda}^{\text{prox}}(\bar{x}) := \text{cone}\left(\left(P_A^{-1}(\bar{x}) \cap \Lambda\right) - \bar{x}\right).$$

When $\Lambda = \mathbb{E}$, it coincides with the proximal normal cone (1).

For $\varepsilon \geq 0$ and $\delta > 0$, a set A is (ε, δ)-*regular* relative to Λ at $\bar{x} \in A$ [13, Definition 2.9] if for all $x \in \mathbb{B}_\delta(\bar{x})$, $a \in A \cap \mathbb{B}_\delta(\bar{x})$ and $v \in N_{A|\Lambda}^{\text{prox}}(a)$,

$$\langle x - a, v \rangle \leq \varepsilon \, \|x - a\| \, \|v\| \, .$$

When $\Lambda = \mathbb{E}$, the quantifier "relative to" is dropped.

For a set-valued operator $T : \mathbb{E} \rightrightarrows \mathbb{E}$, its *fixed point set* is defined by $\mathsf{Fix}\ T := \{x \in \mathbb{E} \mid x \in Tx\}$. For a number $\lambda \in [0, 1]$, we denote the λ-*reflector* of T by $R_{T,\lambda} := (1 + \lambda)T - \lambda \, \mathrm{Id}$. A frequently used example in this paper corresponds to T being a projector.

In the context of convergence analysis of Picard iterations, the following generalization of the Fejér monotonicity of sequences appears frequently, see, for example, the book [4] or the paper [39] for the terminology.

Definition 1 (*Linear monotonicity*) The sequence (x_k) is *linearly monotone* with respect to a set $S \subset \mathbb{E}$ with rate $c \in [0, 1]$ if

$$\mathrm{dist}\, (x_{k+1}, S) \leq c \, \mathrm{dist}\, (x_k, S) \quad \forall k \in \mathbb{N}.$$

Our analysis follows the abstract analysis program proposed in [38] which requires the two key components of the convergence: *almost averagedness* and *metric subregularity*.

Definition 2 (*Almost nonexpansive/averaging mappings*) [38] Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$ and $U \subset \mathbb{E}$.

(i) T is *pointwise almost nonexpansive* at y on U with violation $\varepsilon \geq 0$ if for all $x \in U$, $x^+ \in Tx$ and $y^+ \in Ty$,

$$\left\| x^+ - y^+ \right\| \leq \sqrt{1 + \varepsilon} \, \|x - y\| \, .$$

(ii) T is *pointwise almost averaging* at y on U with violation $\varepsilon \geq 0$ and averaging constant $\alpha > 0$ if for all $x \in U$, $x^+ \in Tx$ and $y^+ \in Ty$,

$$\left\| x^+ - y^+ \right\|^2 \leq (1 + \varepsilon) \, \|x - y\|^2 - \frac{1 - \alpha}{\alpha} \left\| (x^+ - x) - (y^+ - y) \right\|^2 . \quad (2)$$

When a property holds at all $y \in U$ on U, we simply say that the property holds on U.

From Definition 2, almost nonexpansiveness is actually the almost averaging property with the same violation and averaging constant $\alpha = 1$.

Remark 1 (the range of quantitative constants) In the context of Definition 2, it is natural to consider *violation* $\varepsilon \geq 0$ and *averaging constant* $\alpha \in (0, 1]$. Mathematically, it also makes sense to consider $\varepsilon < 0$ and $\alpha > 1$ provided that the required estimate (2) holds true. Simple examples for the later case are linear contraction mappings. In this paper, averaging constant $\alpha > 1$ will frequently be involved implicitly in intermediate

steps of our analysis without any contradiction or confusion. This is the reason why in Definition 2 (ii) we considered $\alpha > 0$ instead of $\alpha \in (0, 1]$ as in [38, Definition 2.2].

It is worth noting that if the iteration $x_{k+1} \in Tx_k$ is linearly monotone with respect to Fix T with rate $c \in (0, 1)$ and T is almost averaging on some neighborhood of Fix T with averaging constant $\alpha \in (0, 1]$, then (x_k) converges R-linearly to a fixed point of T [39, Proposition 3.5].

We next prove a fundamental preliminary result for our analysis regarding almost averaging mappings.

Lemma 1 *Let $T : \mathbb{E} \rightrightarrows \mathbb{E}$, $U \subset \mathbb{E}$, $\lambda \in [0, 1]$, $\varepsilon \geq 0$ and $\alpha > 0$. The following two statements are equivalent.*

(i) T is almost averaging on U with violation ε and averaging constant α.
(ii) The λ-reflector of T, $R_{T,\lambda} = (1 + \lambda)T - \lambda\,\mathrm{Id}$, is almost averaging on U with violation $(1 + \lambda)\varepsilon$ and averaging constant $(1 + \lambda)\alpha$.

Proof Take any $x, y \in U$, $x^+ \in Tx$, $y^+ \in Ty$, $\tilde{x} = (1 + \lambda)x^+ - \lambda x \in R_{T,\lambda}x$ and $\tilde{y} = (1 + \lambda)y^+ - \lambda y \in R_{T,\lambda}y$. We have by definition of $R_{T,\lambda}$ and [4, Corollary 2.14] that

$$\|\tilde{x} - \tilde{y}\|^2 = \left\|(1 + \lambda)(x^+ - y^+) - \lambda(x - y)\right\|^2$$
$$= (1 + \lambda)\left\|x^+ - y^+\right\|^2 - \lambda\|x - y\|^2 + \lambda(1 + \lambda)\left\|(x^+ - x) - (y^+ - y)\right\|^2. \tag{3}$$

We also note that

$$\|(\tilde{x} - x) - (\tilde{y} - y)\| = (1 + \lambda)\left\|(x^+ - x) - (y^+ - y)\right\|. \tag{4}$$

(i) $\Rightarrow$ (ii). Suppose that T is almost averaging on U with violation ε and averaging constant α. Substituting (2) into (3) and using (4), we obtain that

$$\|\tilde{x} - \tilde{y}\|^2$$
$$\leq (1 + (1 + \lambda)\varepsilon)\|x - y\|^2 - (1 + \lambda)\left(\frac{1 - \alpha}{\alpha} - \lambda\right)\left\|(x^+ - x) - (y^+ - y)\right\|^2$$
$$= (1 + (1 + \lambda)\varepsilon)\|x - y\|^2 - \frac{\frac{1-\alpha}{\alpha} - \lambda}{1 + \lambda}\|(\tilde{x} - x) - (\tilde{y} - y)\|^2$$
$$= (1 + (1 + \lambda)\varepsilon)\|x - y\|^2 - \frac{1 - (1 + \lambda)\alpha}{(1 + \lambda)\alpha}\|(\tilde{x} - x) - (\tilde{y} - y)\|^2, \tag{5}$$

which means that $R_{T,\lambda}$ is almost averaging on U with violation $(1+\lambda)\varepsilon$ and averaging constant $(1 + \lambda)\alpha$.

(ii) $\Rightarrow$ (i). Suppose that $R_{T,\lambda}$ is almost averaging on U with violation $(1 + \lambda)\varepsilon$ and averaging constant $(1+\lambda)\alpha$, that is, the inequality (5) is satisfied. Substituting (3) into

(5) and using (4), we obtain

$$(1+\lambda)\left\|x^+ - y^+\right\|^2 - \lambda\|x - y\|^2 + \lambda(1+\lambda)\left\|(x^+ - x) - (y^+ - y)\right\|^2$$
$$\leq (1 + (1+\lambda)\varepsilon)\|x - y\|^2 - (1+\lambda)\left(\frac{1-\alpha}{\alpha} - \lambda\right)\left\|(x^+ - x) - (y^+ - y)\right\|^2.$$

Equivalently,

$$\left\|x^+ - y^+\right\|^2 \leq (1+\varepsilon)\|x - y\|^2 - \frac{1-\alpha}{\alpha}\left\|(x^+ - x) - (y^+ - y)\right\|^2.$$

Hence T is almost averaging on U with violation ε and averaging constant α and the proof is complete. □

Lemma 1 generalizes [13, Lemma 2.4] where the result was proved for $\alpha = 1/2$ and $\lambda = 1$.

The next lemma recalls facts regarding the almost averagedness of projectors and reflectors associated with regular sets.

Lemma 2 *Let $A \subset \mathbb{E}$ be closed and (ε, δ)-regular at $\bar{x} \in A$ and define*

$$U := \{x \in \mathbb{E} \mid P_A x \subset \mathbb{B}_\delta(\bar{x})\}.$$

(i) The projector P_A is pointwise almost nonexpansive on U at every point $z \in A \cap \mathbb{B}_\delta(\bar{x})$ with violation $2\varepsilon + \varepsilon^2$.
(ii) The projector P_A is pointwise almost averaging on U at every point $z \in A \cap \mathbb{B}_\delta(\bar{x})$ with violation $2\varepsilon + 2\varepsilon^2$ and averaging constant $1/2$.
(iii) The λ-reflector $R_{P_A,\lambda}$ is pointwise almost averaging on U at every point $z \in A \cap \mathbb{B}_\delta(\bar{x})$ with violation $(1+\lambda)(2\varepsilon + 2\varepsilon^2)$ and averaging constant $\frac{1+\lambda}{2}$.

Proof Statements (i) and (ii) can be found in [13, Theorem 2.14] or [38, Theorem 3.1 (i) & (iii)]. Statement (iii) follows from (ii) and Lemma 1 applied to $T = P_A$ and $\alpha = 1/2$. □

The following concept of *metric subregularity with functional modulus* has played a central role, explicitly or implicitly, in the convergence analysis of Picard iterations [1,13,38,39]. Recall that a function $\mu : [0, \infty) \to [0, \infty)$ is a *gauge function* if μ is continuous and strictly increasing and $\mu(0) = 0$.

Definition 3 (*Metric subregularity with functional modulus*) A mapping $F : \mathbb{E} \rightrightarrows \mathbb{E}$ is *metrically subregular with gauge μ on $U \subset \mathbb{E}$ for y relative to $\Lambda \subset \mathbb{E}$* if

$$\mu\left(\operatorname{dist}\left(x, F^{-1}(y) \cap \Lambda\right)\right) \leq \operatorname{dist}(y, F(x)) \quad \forall x \in U \cap \Lambda.$$

When μ is a linear function, that is $\mu(t) = \kappa t, \forall t \in [0, \infty)$, one says "with constant κ" instead of "with gauge $\mu = \kappa \operatorname{Id}$". When $\Lambda = \mathbb{E}$, the quantifier "relative to" is dropped.

Metric subregularity has many important applications in variational analysis and mathematical optimization, see the monographs and papers [11,15–18,20,21,25,40, 44]. For the discussion of metric subregularity in connection with subtransversality of collections of sets, we refer the reader to [23,24,29,30].

The next theorem serves as the basic template for the quantitative convergence analysis of fixed point iterations. By the notation $T : \Lambda \rightrightarrows \Lambda$ where Λ is a subset of $\mathbb{E}$, we mean that $T : \mathbb{E} \rightrightarrows \mathbb{E}$ and $Tx \subset \Lambda$ for all $x \in \Lambda$. This simplification of notation should not lead to any confusion if one keeps in mind that there may exist fixed points of T that are not in Λ. For the importance of the use of Λ in isolating the desirable fixed point, we refer the reader to [1, Example 1.8]. In the following, ri Λ denotes the *relative interior* of Λ.

Theorem 1 [38, Theorem 2.1] *Let* $T : \Lambda \rightrightarrows \Lambda$ *for* $\Lambda \subset \mathbb{E}$ *and let* $S \subset$ ri Λ *be closed and nonempty such that* $Ty \subset$ Fix $T \cap S$ *for all* $y \in S$. *Let* $\mathcal{O}$ *be a neighborhood of* S *such that* $\mathcal{O} \cap \Lambda \subset$ ri Λ. *Suppose that*

(a) T *is pointwise almost averaging at all points* $y \in S$ *with violation* ε *and averaging constant* $\alpha \in (0, 1)$ *on* $\mathcal{O} \cap \Lambda$, *and*
(b) *there exists a neighborhood* $\mathcal{V}$ *of* Fix $T \cap S$ *and a constant* $\kappa > 0$ *such that for all* $y \in S$, $y^+ \in Ty$ *and all* $x^+ \in Tx$ *the estimate*

$$\kappa \, dist\,(x, S) \le \left\| (x - x^+) - (y - y^+) \right\| \tag{6}$$

holds whenever $x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$.

Then for all $x^+ \in Tx$

$$dist\,(x^+, \text{Fix } T \cap S) \le \sqrt{1 + \varepsilon - \frac{(1-\alpha)\kappa^2}{\alpha}} \, dist\,(x, S)$$

whenever $x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$.

In particular, if $\kappa > \sqrt{\frac{\varepsilon\alpha}{1-\alpha}}$, *then for any initial point* $x_0 \in \mathcal{O} \cap \Lambda$ *the iteration* $x_{k+1} \in Tx_k$ *satisfies*

$$dist\,(x_{k+1}, \text{Fix } T \cap S) \le c^k \, dist\,(x_0, S)$$

with $c := \sqrt{1 + \varepsilon - \frac{(1-\alpha)\kappa^2}{\alpha}} < 1$ *for all* k *such that* $x_j \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$ *for* $j = 1, 2, \ldots, k$.

Remark 2 [38, p. 13] In the case of $S =$ Fix T condition (6) reduces to metric subregularity of the mapping $F := T -$ Id for 0 on the annular set $(\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda)$, that is
$$\kappa \, dist\,(x, F^{-1}(0)) \le dist\,(0, F(x)) \quad \forall x \in (\mathcal{O} \cap \Lambda) \setminus (\mathcal{V} \cap \Lambda).$$

The inequality $\kappa > \sqrt{\frac{\varepsilon\alpha}{1-\alpha}}$ then states that the constant of metric subregularity κ is sufficiently large relative to the violation of the averaging property of T to guarantee linear progression of the iterates through that annular region.

For a comprehensive discussion on the roles of S and Λ in the analysis program of Theorem 1, we would like to refer the reader to the paper [38].

For the sake of simplification in terms of presentation, we have chosen to reduce the number of technical constants appearing in the analysis. It would be obviously analogous to formulate more theoretically general results by using more technical constants in appropriate places.

3 T_λ as a fixed point operator

We consider the problem of finding a fixed point of the operator

$$T_\lambda := T_1 \left((1 + \lambda)T_2 - \lambda \operatorname{Id} \right) - \lambda \left(T_2 - \operatorname{Id} \right), \tag{7}$$

where $\lambda \in [0, 1]$ and $T_i : \mathbb{E} \rightrightarrows \mathbb{E}$ $(i = 1, 2)$ are assumed to be easily computed.

Examples of T_λ include the backward-backward and the DR algorithms [8,10,34, 36,43] for solving the structured optimization problem

$$\underset{x \in \mathbb{E}}{\text{minimize}} \; f_1(x) + f_2(x)$$

under different assumptions on the functions f_i $(i = 1, 2)$. Indeed, when T_i are the prox mappings of f_i with parameters $\tau_i > 0$, then T_λ with $\lambda = 0$ and 1 takes the form $T_\lambda = \operatorname{prox}_{\tau_1, f_1} \circ \operatorname{prox}_{\tau_2, f_2}$, and $T_\lambda = \operatorname{prox}_{\tau_1, f_1} \left(2\operatorname{prox}_{\tau_2, f_2} - \operatorname{Id} \right) - \operatorname{prox}_{\tau_2, f_2} + \operatorname{Id}$, respectively.

We first characterize the set of fixed points of T_λ via those of the constituent operators T_i $(i = 1, 2)$.

Proposition 1 *Let $T_1, T_2 : \mathbb{E} \rightrightarrows \mathbb{E}$, $\lambda \in [0, 1]$ and consider T_λ defined at (7). The following statements hold true.*

(i) $(1 + \lambda)T_\lambda - \lambda \operatorname{Id} = ((1 + \lambda)T_1 - \lambda \operatorname{Id}) \circ ((1 + \lambda)T_2 - \lambda \operatorname{Id})$.
   As a consequence,

$$\operatorname{Fix} T_\lambda = \operatorname{Fix} \left((1 + \lambda)T_1 - \lambda \operatorname{Id} \right) \circ \left((1 + \lambda)T_2 - \lambda \operatorname{Id} \right).$$

(ii) Suppose that $T_1 = P_A$ is the projector on an affine set A and T_2 is single-valued. Then

$$\begin{aligned}
\operatorname{Fix} T_\lambda &= \{ x \in \mathbb{E} \mid P_A x = \lambda T_2 x + (1 - \lambda)x \} \\
&\subset \{ x \in \mathbb{E} \mid P_A x = P_A T_2 x \}.
\end{aligned} \tag{8}$$

Proof (i). We have by the construction of T_λ that

$$\begin{aligned}
(1 + \lambda)T_\lambda - \lambda \operatorname{Id} &= (1 + \lambda) \left(T_1 \left((1 + \lambda)T_2 - \lambda \operatorname{Id} \right) - \lambda (T_2 - \operatorname{Id}) \right) - \lambda \operatorname{Id} \\
&= (1 + \lambda)T_1 \left((1 + \lambda)T_2 - \lambda \operatorname{Id} \right) - \lambda \left[(1 + \lambda)T_2 - \lambda \operatorname{Id} \right] \\
&= \left((1 + \lambda)T_1 - \lambda \operatorname{Id} \right) \circ \left((1 + \lambda)T_2 - \lambda \operatorname{Id} \right).
\end{aligned}$$

(ii). We first take an arbitrary $x \in \mathsf{Fix}\ T_\lambda$ and prove that

$$P_A x = P_A T_2 x = \lambda T_2 x + (1 - \lambda)x.$$

Indeed, from $x = T_\lambda x$, we get

$$
\begin{aligned}
x &= P_A \left((1 + \lambda)T_2 x - \lambda x\right) - \lambda(T_2 x - x) \\
&\Leftrightarrow \lambda T_2 x + (1 - \lambda)x = P_A \left((1 + \lambda)T_2 x - \lambda x\right).
\end{aligned}
\tag{9}
$$

In particular, $\lambda T_2 x + (1 - \lambda)x \in A$. Thus by equality (9) and the assumption that P_A is affine, we have

$$
\begin{aligned}
P_A \left(\lambda T_2 x + (1 - \lambda)x\right) &= P_A \left((1 + \lambda)T_2 x - \lambda x\right) \\
\Leftrightarrow \lambda P_A T_2 x + (1 - \lambda)P_A x &= (1 + \lambda)P_A T_2 x - \lambda P_A x \\
\Leftrightarrow P_A x &= P_A T_2 x.
\end{aligned}
\tag{10}
$$

Substituting (10) into (9) also yields

$$
\begin{aligned}
\lambda T_2 x + (1 - \lambda)x &= (1 + \lambda)P_A T_2 x - \lambda P_A x \\
&= (1 + \lambda)P_A x - \lambda P_A x = P_A x.
\end{aligned}
$$

Finally, let us take an arbitrary x satisfying $P_A x = \lambda T_2 x + (1 - \lambda)x$ and prove that $x \in \mathsf{Fix}\ T_\lambda$. Indeed, we note that $\lambda T_2 x + (1 - \lambda)x \in A$. Since P_A is affine, one can easily check (10) and then (9), which is equivalent to $x \in \mathsf{Fix}\ T_\lambda$. The proof is complete. □

The inclusion (8) in Proposition 1 can be strict as shown in the next example.

Example 1 Let us consider $\mathbb{E} = \mathbb{R}^2$, the set $A = \left\{(x_1, x_2) \in \mathbb{R}^2 \mid x_1 = 0\right\}$ and the two operators $T_1 = P_A$ and $T_2 x = \frac{1}{2}x$ ($\forall x \in \mathbb{R}^2$). Then for any point $x = (x_1, 0)$ with $x_1 \neq 0$, we have $P_A x = P_A T_2 x = (0, 0)$ but $P_A x = (0, 0) \neq (1 - \lambda/2)x = \lambda T_2 x + (1 - \lambda)x$, that is $x \notin \mathsf{Fix}\ T_\lambda$.

The next proposition shows that the almost averagedness of T_λ naturally inherits from that of T_1 and T_2 via Krasnoselski–Mann relaxations.

Proposition 2 (Almost averagedness of T_λ) *Let* $\lambda \in [0, 1]$, T_i *be almost averaging on* $U_i \subset \mathbb{E}$ *with violation* $\varepsilon_i \geq 0$ *and averaging constant* $\alpha_i > 0$ ($i = 1, 2$) *and define the set*

$$U := \{x \in U_2 \mid R_{T_2, \lambda} x \subset U_1\}.$$

Then T_λ *is almost averaging on* U *with violation* $\varepsilon = \varepsilon_1 + \varepsilon_2 + (1 + \lambda)\varepsilon_1 \varepsilon_2$ *and averaging constant* $\alpha = \frac{2 \max\{\alpha_1, \alpha_2\}}{1 + (1 + \lambda) \max\{\alpha_1, \alpha_2\}}$.

Proof By the implication (i) $\Rightarrow$ (ii) of Lemma 1, the operators $R_{T_i,\lambda} = (1+\lambda)T_i - \lambda\,\mathrm{Id}$ are almost averaging on U_i with violation $(1+\lambda)\varepsilon_i$ and averaging constant $(1+\lambda)\alpha_i$ ($i = 1, 2$). Then thanks to [38, Proposition 2.4 (iii)], the operator $T := R_{T_1,\lambda}R_{T_2,\lambda}$ is almost averaging on U with violation $(1+\lambda)\left(\varepsilon_1 + \varepsilon_2 + (1+\lambda)\varepsilon_1\varepsilon_2\right)$ and averaging constant $\frac{2(1+\lambda)\max\{\alpha_1,\alpha_2\}}{1+(1+\lambda)\max\{\alpha_1,\alpha_2\}}$. Note that $T_\lambda = (1+\lambda)T - \lambda\,\mathrm{Id}$ by Proposition 1. We have by the implication (ii) $\Rightarrow$ (i) of Lemma 1 that T_λ is almost averaging on U with violation $\varepsilon = \varepsilon_1 + \varepsilon_2 + (1+\lambda)\varepsilon_1\varepsilon_2$ and averaging constant $\alpha = \frac{2\max\{\alpha_1,\alpha_2\}}{1+(1+\lambda)\max\{\alpha_1,\alpha_2\}}$ as claimed. $\qquad\square$

We next discuss convergence of T_λ based on the abstract results established in [38]. Our agenda is to verify the assumptions of Theorem 1. To simplify the exposure in terms of presentation, we have chosen to state the results corresponding to $S = \mathsf{Fix}\ T_\lambda$ and $\Lambda = \mathbb{E}$ in Theorem 1. In the sequel, we will denote, for a nonnegative real ρ,

$$S_\rho := \mathsf{Fix}\ T_\lambda + \rho\mathbb{B}.$$

Theorem 2 (Convergence of algorithm T_λ with metric subregularity) *Let T_λ be defined at* (7), $\delta > 0$ *and* $\gamma \in (0, 1)$. *Suppose that for each* $n \in \mathbb{N}$, *the following conditions are satisfied.*

(i) *T_2 is almost averaging on $S_{\gamma^n\delta}$ with violation $\varepsilon_{2,n} \geq 0$ and averaging constant $\alpha_{2,n} \in (0, 1)$, and T_1 is almost averaging on the set $S_{\gamma^n\delta} \cup R_{T_2,\lambda}\left(S_{\gamma^n\delta}\right)$ with violation $\varepsilon_{1,n} \geq 0$ and averaging constant $\alpha_{1,n} \in (0, 1)$.*

(ii) *The mapping $T_\lambda - \mathrm{Id}$ is metrically subregular on $D_n := S_{\gamma^n\delta} \setminus S_{\gamma^{n+1}\delta}$ for 0 with gauge μ_n satisfying*

$$\inf_{x\in D_n} \frac{\mu_n\left(\mathrm{dist}\left(x, \mathsf{Fix}\ T_\lambda\right)\right)}{\mathrm{dist}\left(x, \mathsf{Fix}\ T_\lambda\right)} \geq \kappa_n > \sqrt{\frac{\alpha_n\varepsilon_n}{1 - \alpha_n}}, \tag{11}$$

where $\varepsilon_n := \varepsilon_{1,n} + \varepsilon_{2,n} + (1+\lambda)\varepsilon_{1,n}\varepsilon_{2,n}$ and $\alpha_n := \frac{2\max\{\alpha_{1,n},\alpha_{2,n}\}}{1+(1+\lambda)\max\{\alpha_{1,n},\alpha_{2,n}\}}$.

Then all iterations $x_{k+1} \in T_\lambda x_k$ starting in S_δ satisfy

$$\mathrm{dist}\left(x_k, \mathsf{Fix}\ T_\lambda\right) \to 0 \tag{12}$$

and

$$\mathrm{dist}\left(x_{k+1}, \mathsf{Fix}\ T_\lambda\right) \leq c_n\,\mathrm{dist}\left(x_k, \mathsf{Fix}\ T_\lambda\right) \quad \forall x_k \in D_n, \tag{13}$$

where $c_n := \sqrt{1 + \varepsilon_n - \frac{(1-\alpha_n)\kappa_n^2}{\alpha_n}} < 1$.

In particular, if $\left(\frac{(1-\alpha_n)\kappa_n^2}{\alpha_n} - \varepsilon_n\right)$ is bounded from below by some $\tau > 0$ for all n sufficiently large, then the convergence (12) *is R-linear with rate at most $\sqrt{1 - \tau}$.*

Proof For each $n \in \mathbb{N}$, we verify the assumptions of Theorem 1 for $\mathcal{O} = S_{\gamma^n\delta}$, $\mathcal{V} = S_{\gamma^{n+1}\delta}$ and $D_n = \mathcal{O} \setminus \mathcal{V} = S_{\gamma^n\delta} \setminus S_{\gamma^{n+1}\delta}$. Under assumption (i) of Theorem 2, Proposition 2 ensures that T_λ is almost averaging on $S_{\gamma^n\delta}$ with violation ε_n and averaging constant α_n. In other words, condition (a) of Theorem 1 is satisfied with

$\varepsilon = \varepsilon_n$ and $\alpha = \alpha_n$. Assumption (ii) of Theorem 2 also fulfills condition (b) of Theorem 1 with $\kappa = \kappa_n$ in view of Remark 2. Theorem 1 then yields the conclusion of Theorem 2 after a straightforward care of the involving quantitative constants.   □

The first inequality in (11) essentially says that the gauge function μ_n can be bounded from below by a linear function on the reference interval.

Remark 3 In Theorem 2, the fundamental goal of formulating assumption (i) on the set $S_{\gamma^n \delta}$ and assumption (ii) on the set D_n is that one can characterize sublinear convergence of an iteration on S_δ via linear progression of its iterates through each of the annular set D_n. This idea is based on the fact that for larger n, the almost averaging property of T_λ on $S_{\gamma^n \delta}$ is always improved but the metric subregularity on D_n may get worse, however, if the corresponding quantitative constants still satisfy condition (11), then convergence is guaranteed. For an illustrative example, we refer the reader to [38, Example 2.4].

4 Application to feasibility

We consider algorithm T_λ for solving feasibility problem involving two closed sets $A, B \subset \mathbb{E}$,

$$
\begin{aligned}
x^+ \in T_\lambda x &= P_A \left((1 + \lambda) P_B x - \lambda x \right) - \lambda \left(P_B x - x \right) \\
&= P_A R_{P_B, \lambda}(x) - \lambda \left(P_B x - x \right).
\end{aligned}
\tag{14}
$$

Note that T_λ with $\lambda = 0$ and 1 corresponds to the alternating projections $P_A P_B$ and the DR method $\frac{1}{2}(R_A \circ R_B + \mathrm{Id})$, respectively.

It is worth recalling that feasibility problem for $m \geq 2$ sets can be reformulated as a feasibility problem for two constructed sets on the product space $\mathbb{E}^m$ with one of the later sets is a linear subspace, and the regularity properties in terms of both individual sets and collections of sets of the later sets are inherited from those of the former ones [3,32].

When A is an affine set, then the projector P_A is affine and T_λ is a convex combination of the alternating projection and the DR methods since

$$
\begin{aligned}
T_\lambda x &= P_A \left((1 - \lambda) P_B x + \lambda (2 P_B x - x) \right) - \lambda \left(P_B x - x \right) \\
&= (1 - \lambda) P_A P_B x + \lambda \left(x + P_A (2 P_B x - x) - P_B x \right) \\
&= (1 - \lambda) T_0(x) + \lambda T_1(x).
\end{aligned}
$$

In this case, we establish convergence results for all convex combinations of the alternating projection and the DR methods. To our best awareness, this kind of results seems to be new.

Recall that when applied to inconsistent feasibility problems the DR operator has no fixed points. We next show that the set of fixed points of T_λ with $\lambda \in [0, 1)$ for convex inconsistent feasibility problems is nonempty. This result follows the lines of [36, Lemma 2.1] where the fixed point set of the RAAR operator is characterized.

Proposition 3 (Fixed points of T_λ for convex inconsistent feasibility) *For closed convex sets A, $B \subset \mathbb{E}$, let $G = \overline{B - A}$, $g = P_G 0$, $E = A \cap (B - g)$ and $F = (A + g) \cap B$. Then*

$$\mathsf{Fix}\, T_\lambda = E - \frac{\lambda}{1 - \lambda} g \quad \forall \lambda \in [0, 1).$$

Proof We first show that $E - \frac{\lambda}{1-\lambda} g \subset \mathsf{Fix}\, T_\lambda$. Pick any $e \in E$ and denote $f = e + g \in F$ as definitions of E and F. We are checking that

$$x := e - \frac{\lambda}{1 - \lambda} g \in \mathsf{Fix}\, T_\lambda.$$

Since $x = f - \frac{1}{1-\lambda} g$ and $-g \in N_B(f)$, we get $P_B x = f$.

Analogously, since $g \in N_A(e)$ and

$$(1 + \lambda) P_B x - \lambda x = (1 + \lambda) f - \lambda x = e + \frac{1}{1 - \lambda} g,$$

we have $P_A((1 + \lambda) P_B x - \lambda x) = e$.

Hence,

$$x - T_\lambda x = x - P_A \left((1 + \lambda) P_B x - \lambda x\right) + \lambda \left(P_B x - x\right)$$
$$= x - e + \lambda (f - x) = 0.$$

That is $x \in \mathsf{Fix}\, T_\lambda$.

We next show that $\mathsf{Fix}\, T_\lambda \subset E - \frac{\lambda}{1-\lambda} g$. Pick any $x \in \mathsf{Fix}\, T_\lambda$. Let $f = P_B x$ and $y = x - f$. Thanks to $x \in \mathsf{Fix}\, T_\lambda$ and the definition of T_λ,

$$P_A((1 + \lambda) P_B x - \lambda x) = \lambda(P_B x - x) + x$$
$$= -\lambda y + y + f = f + (1 - \lambda) y. \tag{15}$$

Now, for any $a \in A$, since A is closed and convex, we have

$$0 \geq \langle a - P_A((1 + \lambda) P_B x - \lambda x), (1 + \lambda) P_B x - \lambda x - P_A((1 + \lambda) P_B x - \lambda x) \rangle$$
$$= \langle a - (f + (1 - \lambda) y), (1 + \lambda) f - \lambda x - (f + (1 - \lambda) y) \rangle$$
$$= \langle a - f - (1 - \lambda) y, -y \rangle = \langle -a + f, y \rangle + (1 - \lambda) \|y\|^2.$$

On the other hand, for any $b \in B$, since B is closed and convex, we have

$$\langle b - f, y \rangle = \langle b - f, x - f \rangle = \langle b - P_B x, x - P_B x \rangle \leq 0.$$

Combining the last two inequalities yields

$$\langle b - a, y \rangle \leq -(1 - \lambda) \|y\|^2 \leq 0 \quad \forall a \in A, \forall b \in B.$$

Take a sequence (a_n) in A and a sequence (b_n) in B such that $g_n := b_n - a_n \to g$. Then

$$\langle g_n, y \rangle \leq -(1 - \lambda) \|y\|^2 \leq 0 \quad \forall n. \tag{16}$$

Taking the limit and using the Cauchy–Schwarz inequality yields

$$\|y\| \leq \frac{1}{1 - \lambda} \|g\|.$$

Conversely, by (15) with noting that $f \in B$ and $P_A((1 + \lambda)P_B x - \lambda x) \in A$,

$$\|y\| = \frac{1}{1 - \lambda} \|f - P_A((1 + \lambda)P_B x - \lambda x)\| \geq \frac{1}{1 - \lambda} \|g\|.$$

Hence $\|y\| = \frac{1}{1-\lambda} \|g\|$, and taking the limit in (16), which yields $y = -\frac{1}{1-\lambda} g$. Since $f \in B$ and $f - g = f + (1 - \lambda)y = P_A((1 + \lambda)P_B x - \lambda x) \in A$, we have $f - g \in A \cap (B - g) = E$ and, therefore,

$$x = f + y = f - \frac{1}{1 - \lambda} g = f - g - \frac{\lambda}{1 - \lambda} g \in E - \frac{\lambda}{1 - \lambda} g.$$

$\square$

We next discuss the two key ingredients for convergence of algorithm T_λ applied to feasibility problems: 1) almost averagedness of T_λ, and 2) metric subregularity of $T_\lambda - \text{Id}$. The two properties will be deduced from the (ε, δ)-regularity of the individual sets and the transversality of the collection of sets, respectively.

The next proposition shows averagedness of T_λ applied to feasibility problems involving (ε, δ)-regular sets.

Proposition 4 *Let A and B be (ε, δ)-regular at $\bar{x} \in A \cap B$ and define the set*

$$U := \{x \in \mathbb{E} \mid P_B x \subset \mathbb{B}_\delta(\bar{x}) \text{ and } P_A R_{P_B, \lambda} x \subset \mathbb{B}_\delta(\bar{x})\}. \tag{17}$$

Then T_λ is pointwise almost averaging on U at every point $z \in S := A \cap B \cap \mathbb{B}_\delta(\bar{x})$ with averaging constant $\frac{2}{3+\lambda}$ and violation

$$\tilde{\varepsilon} := 2(2\varepsilon + 2\varepsilon^2) + (1 + \lambda)(2\varepsilon + 2\varepsilon^2)^2. \tag{18}$$

Proof Let us define the two sets

$$U_A := \{y \in \mathbb{E} \mid P_A y \subset \mathbb{B}_\delta(\bar{x})\}, \quad U_B := \{x \in \mathbb{E} \mid P_B x \subset \mathbb{B}_\delta(\bar{x})\}$$

and note that $x \in U$ if and only if $x \in U_B$ and $R_{P_B, \lambda} x \subset U_A$. Thanks to Lemma 2 (iii), $R_{P_A, \lambda}$ and $R_{P_B, \lambda}$ are pointwise almost averaging at every point $z \in S$ with violation $(1 + \lambda)(2\varepsilon + 2\varepsilon^2)$ and averaging constant $\frac{1+\lambda}{2}$ on U_A and U_B, respectively. Then due to [38, Proposition 2.4 (iii)], the operator $T := R_{P_A, \lambda} R_{P_B, \lambda}$ is pointwise almost

averaging on U at every point $z \in S$ with averaging constant $\frac{2(1+\lambda)}{3+\lambda}$ and violation $(1 + \lambda)\tilde{\varepsilon}$, where $\tilde{\varepsilon}$ is given by (18). Note that $T_\lambda = (1 + \lambda)T - \lambda\,\mathrm{Id}$ by Proposition 1. Thanks to Lemma 1, T_λ is pointwise almost averaging on U at every point $z \in S$ with violation $\tilde{\varepsilon}$ and averaging constant $\frac{2}{3+\lambda}$ as claimed. □

Remark 4 It follows from Lemma 2 (i) & (iii) that the set U defined by (17) contains at least the ball $\mathbb{B}_{\delta'}(\bar{x})$, where

$$\delta' := \frac{\delta}{2(1+\varepsilon)\sqrt{1 + (1+\lambda)(2\varepsilon + 2\varepsilon^2)}} > 0.$$

We next integrate Proposition 4 into Theorem 2 to obtain convergence of algorithm T_λ for solving consistent feasibility problems involving (ε, δ)-regular sets.

Corollary 1 (Convergence of algorithm T_λ for feasibility) *Consider the algorithm T_λ defined at (14) and suppose that* $\mathsf{Fix}\, T_\lambda = A \cap B \neq \emptyset$*. Denote $S_\rho = \mathsf{Fix}\, T_\lambda + \rho\mathbb{B}$ for a nonnegative real ρ. Suppose that there are $\delta > 0$, $\varepsilon \geq 0$ and $\gamma \in (0, 1)$ such that A and B are (ε, δ')-regular at avery point $z \in A \cap B$, where*

$$\delta' := 2\delta(1+\varepsilon)\sqrt{1 + (1+\lambda)(2\varepsilon + 2\varepsilon^2)},$$

and for each $n \in \mathbb{N}$, the mapping $T_\lambda - \mathrm{Id}$ is metrically subregular on $D_n := S_{\gamma^n\delta} \setminus S_{\gamma^{n+1}\delta}$ for 0 with gauge μ_n satisfying

$$\inf_{x \in D_n} \frac{\mu_n\,(dist\,(x, A \cap B))}{dist\,(x, A \cap B)} \geq \kappa_n > \sqrt{\frac{2\tilde{\varepsilon}}{1+\lambda}},$$

where $\tilde{\varepsilon}$ is given at (18).

Then all iterations $x_{k+1} \in T_\lambda x_k$ starting in S_δ satisfy (12) and (13) with $c_n := \sqrt{1 + \tilde{\varepsilon} - \frac{(1+\lambda)\kappa_n^2}{2}} < 1$.

In particular, if (κ_n) is bounded from below by some $\kappa > \sqrt{\frac{2\tilde{\varepsilon}}{1+\lambda}}$ for all n sufficiently large, then (x_k) eventually converges R-linearly to a point in $A \cap B$ with rate at most $\sqrt{1 + \tilde{\varepsilon} - \frac{(1+\lambda)\kappa^2}{2}} < 1$.

Proof Let any $x \in D_n$, for some $n \in \mathbb{N}$, $x^+ \in T_\lambda x$ and $\bar{x} \in P_{A \cap B}x$. A combination of Proposition 4 and Remark 4 implies that T_λ is pointwise almost averaging on $\mathbb{B}_\delta(\bar{x})$ at every point $z \in A \cap B \cap \mathbb{B}_\delta(\bar{x})$ with violation $\tilde{\varepsilon}$ given by (18) and averaging constant $\frac{2}{3+\lambda}$. In other words, condition (a) of Theorem 1 is satisfied. Condition (b) of Theorem 1 is also fulfilled by the same argument as the one used in Theorem 2. The desired conclusion now follows from Theorem 1. □

In practice, the metric subregularity assumption is often more challenging to be verified than the averaging property. In the concrete example of consistent alternating projections $P_A P_B$, that metric subregularity condition holds true if and only if the collection of sets is subtransversal. We next show that the metric subregularity of

$T_\lambda - $ Id can be deduced from the transversality of the collection of sets $\{A, B\}$. As a result, if the sets are also sufficiently regular, then local linear convergence of the iteration $x_{k+1} \in T_\lambda x_k$ is guaranteed.

We first describe the concept of relative transversality of collections of sets. In the sequel, we set $\Lambda := \text{aff}(A \cup B)$, the smallest affine set in $\mathbb{E}$ containing both A and B.

Assumption 3 The collection $\{A, B\}$ is transversal at $\bar{x} \in A \cap B$ relative to Λ with constant $\bar{\theta} < 1$, that is, for any $\theta \in (\bar{\theta}, 1)$, there exists $\delta > 0$ such that

$$\langle u, v \rangle \geq -\theta \, \|u\| \cdot \|v\|$$

holds for all $a \in A \cap \mathbb{B}_\delta(\bar{x})$, $b \in B \cap \mathbb{B}_\delta(\bar{x})$, $u \in N_{A|\Lambda}^{\text{prox}}(a)$ and $v \in N_{B|\Lambda}^{\text{prox}}(b)$.

Thanks to [22, Theorem 1] and [28, Theorem 1], Assumption 3 also ensures subtransversality of $\{A, B\}$ at $\bar{x}$ relative to Λ with constant at least $\sqrt{\frac{1-\theta}{2}}$ on the neighborhood $\mathbb{B}_\delta(\bar{x})$, that is

$$\sqrt{\frac{1-\theta}{2}} \, \text{dist}\,(x, A \cap B) \leq \max\{\text{dist}\,(x, A), \text{dist}\,(x, B)\} \quad \forall x \in \Lambda \cap \mathbb{B}_\delta(\bar{x}). \quad (19)$$

The next lemma is at the heart of our subsequent discussion.

Lemma 3 *Suppose that Assumption 3 is satisfied. Then for any $\theta \in (\bar{\theta}, 1)$, there exists a number $\delta > 0$ such that for all $x \in \mathbb{B}_\delta(\bar{x})$ and $x^+ \in T_\lambda x$,*

$$\kappa \, \text{dist}\,(x, A \cap B) \leq \|x - x^+\|, \quad (20)$$

where κ is defined by

$$\kappa := \frac{(1-\theta)\sqrt{1+\theta}}{\sqrt{2} \max\left\{1, \lambda + \sqrt{1-\theta^2}\right\}} > 0. \quad (21)$$

Proof For any $\theta \in (\bar{\theta}, 1)$, there is a number $\delta > 0$ satisfying the property described in Assumption 3. Let us set $\delta' = \delta/6$ and show that condition (20) is fulfilled with δ'.

Indeed, let us consider any $x \in \mathbb{B}_{\delta'}(\bar{x})$, $b \in P_B x$, $y = (1+\lambda)b - \lambda x$, $a \in P_A y$ and $x^+ = a - \lambda(b - x) \in T_\lambda x$. From the choice of δ', it is clear that $a, b \in \mathbb{B}_\delta(\bar{x})$. Since $x - b \in N_{B|\Lambda}^{\text{prox}}(b)$ and $y - a \in N_{A|\Lambda}^{\text{prox}}(a)$, Assumption 3 yields that

$$\langle x - b, y - a \rangle \geq -\theta \, \|x - b\| \cdot \|y - a\|. \quad (22)$$

By the definition of T_λ, we have

$$
\begin{aligned}
\left\| x - x^+ \right\|^2 &= \| x - b + y - a \|^2 \\
&= \| x - b \|^2 + \| y - a \|^2 + 2 \langle x - b, y - a \rangle \\
&\geq \| x - b \|^2 + \| y - a \|^2 - 2\theta \, \| x - b \| \cdot \| y - a \| \\
&\geq \left(1 - \theta^2 \right) \| x - b \|^2 = \left(1 - \theta^2 \right) \operatorname{dist}^2(x, B), \tag{23}
\end{aligned}
$$

where the first inequality follows from (22).

We will take care of the two possible cases regarding $\operatorname{dist}(x, A)$ as follows.

Case 1 $\operatorname{dist}(x, A) \leq \left(\lambda + \sqrt{1 - \theta^2} \right) \operatorname{dist}(x, B)$. Thanks to (23) we get

$$
\left\| x - x^+ \right\|^2 \geq \frac{1 - \theta^2}{\left(\lambda + \sqrt{1 - \theta^2} \right)^2} \operatorname{dist}^2(x, A). \tag{24}
$$

Case 2 $\operatorname{dist}(x, A) > \left(\lambda + \sqrt{1 - \theta^2} \right) \operatorname{dist}(x, B)$. By the triangle inequality and the construction of T_λ, we get

$$
\begin{aligned}
\left\| x - x^+ \right\| &\geq \| x - a \| - \left\| a - x^+ \right\| = \| x - a \| - \lambda \| x - b \| \\
&\geq \operatorname{dist}(x, A) - \lambda \operatorname{dist}(x, B) \geq \left(1 - \frac{\lambda}{\lambda + \sqrt{1 - \theta^2}} \right) \operatorname{dist}(x, A). \tag{25}
\end{aligned}
$$

Since

$$
\frac{1 - \theta^2}{\left(\lambda + \sqrt{1 - \theta^2} \right)^2} = \left(1 - \frac{\lambda}{\lambda + \sqrt{1 - \theta^2}} \right)^2,
$$

we always have from (24) and (25) that

$$
\left\| x - x^+ \right\|^2 \geq \frac{1 - \theta^2}{\left(\lambda + \sqrt{1 - \theta^2} \right)^2} \operatorname{dist}^2(x, A). \tag{26}
$$

Combining (23), (26) and (19), we obtain

$$
\begin{aligned}
\left\| x - x^+ \right\|^2 &\geq \frac{1 - \theta^2}{\max \left\{ 1, \left(\lambda + \sqrt{1 - \theta^2} \right)^2 \right\}} \max \left\{ \operatorname{dist}^2(x, A), \operatorname{dist}^2(x, B) \right\} \\
&\geq \frac{(1 - \theta^2)(1 - \theta)}{2 \max \left\{ 1, \left(\lambda + \sqrt{1 - \theta^2} \right)^2 \right\}} \operatorname{dist}^2(x, A \cap B),
\end{aligned}
$$

which yields (20) as claimed. $\square$

In the special case that $\lambda = 1$, Lemma 3 refines [13, Lemma 3.14] and [45, Lemma 4.2] where the result was proved for the DR operator with an additional assumption on regularity of the sets.

The next result is the final preparation for our linear convergence result.

Lemma 4 [45, Proposition 2.11] *Let* $T : \mathbb{E} \rightrightarrows \mathbb{E}$, $S \subset \mathbb{E}$ *be closed and* $\bar{x} \in S$. *Suppose that there are* $\delta > 0$ *and* $c \in [0, 1)$ *such that for all* $x \in \mathbb{B}_\delta(\bar{x})$, $x^+ \in Tx$ *and* $z \in P_S x$,

$$\left\| x^+ - z \right\| \le c \left\| x - z \right\|. \tag{27}$$

Then every iteration $x_{k+1} \in Tx_k$ *starting sufficiently close to* $\bar{x}$ *converges R-linearly to a point* $\tilde{x} \in S \cap \mathbb{B}_\delta(\bar{x})$. *In particular,*

$$\|x_k - \tilde{x}\| \le \frac{\|x_0 - \bar{x}\| (1 + c)}{1 - c} c^k.$$

We are now ready to prove local linear convergence for algorithm T_λ which generalizes the corresponding results established in [13,45] for the DR method.

Theorem 4 (Linear convergence of algorithm T_λ for feasibility) *In addition to Assumption 3, suppose that A and B are* (ε, δ)-*regular at* $\bar{x}$ *with* $\tilde{\varepsilon} < \frac{(1+\lambda)\kappa^2}{2}$, *where* $\tilde{\varepsilon}$ *and* κ *are given by* (18) *and* (21), *respectively. Then every iteration* $x_{k+1} \in T_\lambda x_k$ *starting sufficiently close to* $\bar{x}$ *converges R-linearly to a point in* $A \cap B$.

Proof Assumption 3 ensures the existence of $\delta_1 > 0$ such that Lemma 3 holds true. In view of Proposition 4 and Remark 4, one can find a number $\delta_2 > 0$ such that T_λ is pointwise almost averaging on $\mathbb{B}_{\delta_2}(\bar{x})$ at every point $z \in A \cap B \cap \mathbb{B}_{\delta_2}(\bar{x})$ with violation $\tilde{\varepsilon}$ given by (18) and averaging constant $\frac{2}{3+\lambda}$. Define $\delta' = \min\{\delta_1, \delta_2\} > 0$.

Now let us consider any $x \in \mathbb{B}_{\delta'/2}(\bar{x})$, $x^+ \in T_\lambda x$ and $z \in P_{A \cap B} x$. It is clear that $z \in \mathbb{B}_{\delta'}(\bar{x})$. Proposition 4 and Lemma 3 then respectively yield

$$\left\| x^+ - z \right\|^2 \le (1 + \tilde{\varepsilon}) \|x - z\|^2 - \frac{1 + \lambda}{2} \left\| x - x^+ \right\|^2, \tag{28}$$

$$\left\| x - x^+ \right\|^2 \ge \kappa^2 \operatorname{dist}^2(x, A \cap B) = \kappa^2 \|x - z\|^2, \tag{29}$$

where κ is given by (21).

Substituting (29) into (28), we get

$$\left\| x^+ - z \right\|^2 \le \left(1 + \tilde{\varepsilon} - \frac{(1 + \lambda)\kappa^2}{2} \right) \|x - z\|^2,$$

which yields condition (27) of Lemma 4 and the desired conclusion now follows from this lemma.   □

5 Application to sparse optimization

Our goal in this section is twofold: 1) to illustrate the linear convergence of algorithm T_λ formulated in Theorem 4 via the *sparse optimization problem*, and 2) to demonstrate

a promising performance of the algorithm T_λ in comparison with the RAAR algorithm for this applied problem.

5.1 Sparse optimization

We consider the *sparse optimization problem*

$$\min_{x \in \mathbb{R}^n} \|x\|_0 \quad \text{subject to} \quad Mx = b, \tag{30}$$

where $M \in \mathbb{R}^{m \times n}$ ($m < n$) is a full rank matrix, b is a given vector in $\mathbb{R}^m$, and $\|x\|_0$ is the number of nonzero entries of the vector x. The sparse optimization problem with complex variable is defined analogously by replacing $\mathbb{R}$ by $\mathbb{C}$ everywhere in the above model.

Many strategies for solving (30) have been proposed. We refer the reader to the famous paper by Candès and Tao [9] for solving this problem by using convex relaxations. On the other hand, assuming to have a good guess on the sparsity of the solutions to (30), one can tackle this problem by solving the *sparse feasibility problem* [14] of finding

$$\bar{x} \in A_s \cap B, \tag{31}$$

where $A_s := \{x \in \mathbb{R}^n \mid \|x\|_0 \leq s\}$ and $B := \{x \in \mathbb{R}^n \mid Mx = b\}$.

It is worth mentioning that the initial guess s of the true sparsity is not numerically sensitive with respect to various projection methods, that is, for a relatively wide range of values of s above the true sparsity, projection algorithms perform very much in the same nature. Note also that the approach via sparse feasibility does not require convex relaxations of (30) and thus can avoid the likely expensive increase of dimensionality.

We run the two algorithms T_λ and RAAR to solve (31) and compare their numerical performances. By taking s smaller than the true sparsity, we can also compare their performances for inconsistent feasibility.

Since B is affine, there is the closed algebraic form for the projector P_B,

$$P_B x = x - M^\dagger (Mx - b) \quad \forall x \in \mathbb{R}^n,$$

where $M^\dagger := M^T (MM^T)^{-1}$ is the *Moore–Penrose inverse* of M. We have denoted M^T the transpose matrix of M and taken into account that M is full rank. There is also a closed form for P_{A_s} [6]. For each $x \in \mathbb{R}^n$, let us denote $\mathcal{I}_s(x)$ the set of all s-tubles of indices of s largest in absolute value entries of x. The set $\mathcal{I}_s(x)$ can contain multiple such s-tubles. The projector P_{A_s} can be described as

$$P_{A_s} x = \left\{ z \in \mathbb{R}^n \mid \exists \mathcal{I} \in \mathcal{I}_s(x) \text{ such that } z(k) = \begin{cases} x(k) & \text{if } k \in \mathcal{I}, \\ 0 & \text{else} \end{cases} \right\}.$$

For convenience, we recall the two algorithms in this specific setting

$$RAAR_\beta = \beta \left(P_{A_s}(2P_B - \text{Id}) \right) + (1 - 2\beta)P_B + \beta \, \text{Id},$$
$$T_\lambda = P_{A_s} \left((1 + \lambda)P_B - \lambda \, \text{Id} \right) - \lambda(P_B - \text{Id}).$$

5.2 Convergence analysis

We analyze the convergence of algorithm T_λ for the sparse feasibility problem (31). The next theorem establishes local linear convergence of algorithm T_λ for solving sparse feasibility problems.

Theorem 5 (Linear convergence of algorithm T_λ for sparse feasibility) *Let $\bar{x} = (\bar{x}_i) \in A_s \cap B$ and suppose that s is the sparsity of the solutions to the problem* (30). *Then any iteration $x_{k+1} \in T_\lambda x_k$ starting sufficiently close to $\bar{x}$ converges R-linearly to $\bar{x}$.*

Proof We first show that $\bar{x}$ is an isolated point of $A_s \cap B$. Since s is the sparsity of the solutions to (30), we have that $\|\bar{x}\|_0 = s$ and the set $\mathcal{I}_s(\bar{x})$ contains a unique element, denoted $\mathcal{I}_{\bar{x}}$. Note that $E_{\bar{x}} := \text{span}\{e_i : i \in \mathcal{I}_{\bar{x}}\}$ is the unique s-dimensional space component of A_s containing $\bar{x}$, where $\{e_i : 1 \le i \le n\}$ is the canonical basic of $\mathbb{R}^n$. Let us denote

$$\delta := \min_{i \in \mathcal{I}_{\bar{x}}} |\bar{x}_i| > 0.$$

We claim that

$$A_s \cap \mathbb{B}_\delta(\bar{x}) = E_{\bar{x}} \cap \mathbb{B}_\delta(\bar{x}), \tag{32}$$
$$E_{\bar{x}} \cap B = \{\bar{x}\}. \tag{33}$$

Indeed, for any $x = (x_i) \in A_s \cap \mathbb{B}_\delta(\bar{x})$, we have by definition of δ that $x_i \ne 0$ for all $i \in \mathcal{I}_{\bar{x}}$. Hence $\|x\|_0 = s$ and $x \in E_{\bar{x}} \cap \mathbb{B}_\delta(\bar{x})$. This proves (32).

For (33), it suffices to show the singleton of $E_{\bar{x}} \cap B$ since we already know that $\bar{x} \in E_{\bar{x}} \cap B$. Suppose otherwise that there exists $x = (x_i) \in E_{\bar{x}} \cap B$ with $x_j \ne \bar{x}_j$ for some index j. Since both $E_{\bar{x}}$ and B are affine, the intersection $E_{\bar{x}} \cap B$ contains the line $\{x + t(\bar{x} - x) : t \in \mathbb{R}\}$ passing x and $\bar{x}$. In particular, it contains the point $z := x + \frac{x_j}{x_j - \bar{x}_j}(\bar{x} - x)$. Then we have that $z \in B$ and $\|z\|_0 \le s - 1$ as $z_j = 0$. This contradicts to the assumption that s is the sparsity of the solutions to (30), and hence (33) is proved.

A combination of (32) and (33) then yields

$$A_s \cap B \cap \mathbb{B}_\delta(\bar{x}) = E_{\bar{x}} \cap B \cap \mathbb{B}_\delta(\bar{x}) = \{\bar{x}\}. \tag{34}$$

This means that $\bar{x}$ is an isolated point of $A_s \cap B$ as claimed. Moreover, the equalities in (34) imply that

$$P_{A_s}x = P_{E_{\bar{x}}}x \quad \forall x \in \mathbb{B}_{\delta/2}(\bar{x}).$$

Therefore, for any starting point $x_0 \in \mathbb{B}_{\delta/2}(\bar{x})$, the iteration $x_{k+1} \in T_\lambda x_k$ for solving (31) is identical to that for solving the feasibility problem for the two sets $E_{\bar{x}}$ and B.

Since $E_{\bar{x}}$ and B are two affine subspaces intersecting at the unique point $\bar{x}$ by (33), the collection of sets $\{E_{\bar{x}}, B\}$ is transversal at $\bar{x}$ relative to the affine hull aff $(E_{\bar{x}} \cup B)$. Theorem 4 now can be applied to conclude that the iteration $x_{k+1} \in T_{\lambda}x_k$ converges R-linearly to $\bar{x}$. The proof is complete.   □

It is worth mentioning that the convergence analysis in Theorem 5 is also valid for the RAAR algorithm.

5.3 Numerical experiment

We now set up a toy example as in [9,14] which involves an unknown true object $\bar{x} \in \mathbb{R}^{256^2}$ with $\|\bar{x}\|_0 = 328$ (the sparsity rate is .005). Let b be 1/8 of the measurements of $F(\bar{x})$, the *Fourier transform* of $\bar{x}$, with the *sample indices* denoted $\mathcal{J}$. The *Poisson noise* was added when calculating the measurement b. Note that since $\bar{x}$ is real, $F(\bar{x})$ is *conjugate symmetric*, we indeed have nearly a double number of measurements. In this setting, we have

$$B = \{x \in \mathbb{C}^{256^2} \mid F(x)(k) = b(k), \ \forall k \in \mathcal{J}\},$$

and the two prox operators, respectively, take the forms

$$P_{A_s}x = \left\{ z \in \mathbb{R}^n \mid \exists \mathcal{I} \in \mathcal{I}_s(x) \ \text{such that} \ z(k) = \begin{cases} \mathrm{Re}\,(x(k)) & \text{if } k \in \mathcal{I}, \\ 0 & \text{else} \end{cases} \right\},$$

$$P_B x = F^{-1}(\hat{x}), \quad \text{where} \ \hat{x}(k) = \begin{cases} b(k) & \text{if } k \in \mathcal{J}, \\ F(x)(k) & \text{else,} \end{cases}$$

where $\mathrm{Re}(x(k))$ denotes the real part of the complex number $x(k)$, and F^{-1} is the *inverse Fourier transform*.

The initial point was chosen randomly, and a warm-up procedure with 10 DR iterates was performed before running the two algorithms. The stopping criterion $\|x - x^+\| < 10^{-10}$ was used. We have used the *Matlab ProxToolbox* [37] to run this numerical experiment. The parameters were chosen in such a way that the performance is seemingly optimal for both algorithms. We chose $\beta = .65$ for the RAAR algorithm and $\lambda = .45$ for algorithm T_{λ} in the case of consistent feasibility problem corresponding to $s = 340$, and $\beta = .6$ for the RAAR algorithm and $\lambda = .4$ for algorithm T_{λ} in the case of inconsistent feasibility problem corresponding to $s = 310$.

The *change* of distances between two consecutive iterates is of interest. When linear convergence appears to be the case, it can yield useful information of the convergence rate. Under the assumption that the iterates will remain in the convergence area, one can obtain error bounds for the distance from the current iterate to a nearest solution. We also pay attention to the *gaps* in iterates that in a sense measure the infeasibility at the iterates. If we think feasibility problem as the problem of minimizing the sum of the squares of the distance functions to the sets, then gaps in iterates are the values of that function evaluated at the iterates. For the two algorithms under consideration,
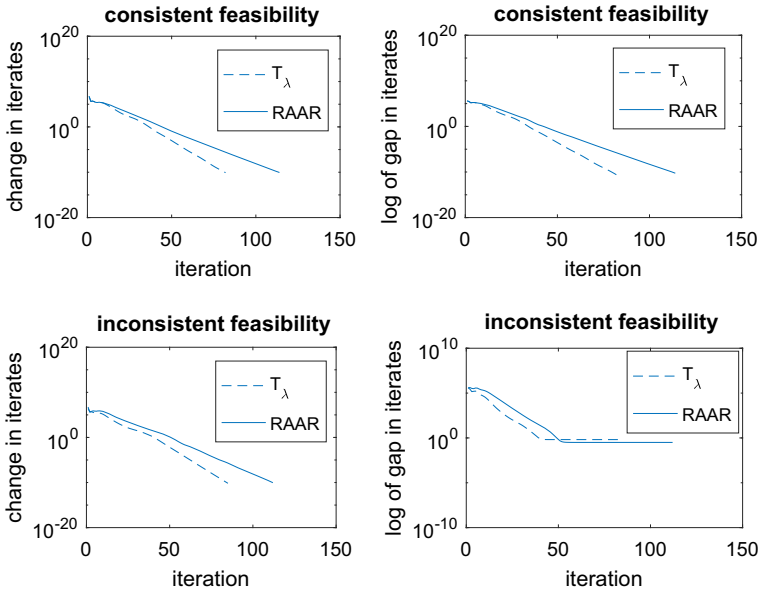
Fig. 1 Performances of the RAAR and T_λ algorithms for sparse feasibility problem: iterate changes in consistent case (top-left), iterate gaps in consistent case (top-right), iterate changes in inconsistent case (bottom-left) and iterate gaps in inconsistent case (bottom-right)

the iterates are themselves not informative but their shadows, by which we mean the projections of the iterates on one of the sets. Hence, the gaps in iterates are calculated for the iterate shadows instead of the iterates themselves.

Figure 1 summarizes the performances of the two algorithms for both consistent and inconsistent sparse feasibility problems. We first emphasize that the algorithms appear to be convergent in both cases of feasibility. For the consistent case, algorithm T_λ appears to perform better than the RAAR algorithm in terms of both the iterate changes and gaps. Also, the CPU time of algorithm T_λ is around 10% less than that of the RAAR algorithm. For the inconsistent case, we have a similar observation except that the iterate gaps for the RAAR algorithm are slightly better (smaller) than those for algorithm T_λ. Extensive numerical experiments in imaging problems illustrating the empirical performance of algorithm T_λ will be the future work.

References

1. Aspelmeier, T., Charitha, C., Luke, D.R.: Local linear convergence of the ADMM/Douglas–Rachford algorithms without strong convexity and application to statistical imaging. SIAM J. Imaging Sci. **9**(2), 842–868 (2016)
2. Attouch, H., Bolte, J., Redont, P., Soubeyran, A.: Proximal alternating minimization and projection methods for nonconvex problems: an approach based on the Kurdyka–Łojasiewicz inequality. Math. Oper. Res. **35**(2), 438–457 (2010)
3. Bauschke, H.H., Borwein, J.M.: On projection algorithms for solving convex feasibility problems. SIAM Rev. **38**(3), 367–426 (1996)
4. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)
5. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and the method of alternating projections: applications. Set-Valued Var. Anal. **21**, 475–501 (2013)
6. Bauschke, H.H., Luke, D.R., Phan, H.M., Wang, X.: Restricted normal cones and sparsity optimization with affine constraints. Found. Comput. Math. **14**, 63–83 (2014)
7. Bolte, J., Sabach, S., Teboulle, M.: Proximal alternating linearized minimization for nonconvex and nonsmooth problems. Math. Program. **146**(1–2), 459–494 (2014)
8. Borwein, J.M., Tam, M.K.: The cyclic Douglas–Rachford method for inconsistent feasibility problems. J. Nonlinear Convex Anal. **16**(4), 537–584 (2015)
9. Candés, E., Tao, T.: Decoding by linear programming. IEEE Trans. Inf. Theory **51**(12), 4203–4215 (2005)
10. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, vol. 49. Springer, Berlin, pp. 185–212 (2011)
11. Dontchev, A.L., Rockafellar, R.T.: Implicit Functions and Solution Mapppings. Srpinger, New York (2014)
12. Drusvyatskiy, D., Ioffe, A.D., Lewis, A.S.: Transversality and alternating projections for nonconvex sets. Found. Comput. Math. **15**(6), 1637–1651 (2015)
13. Hesse, R., Luke, D.R.: Nonconvex notions of regularity and convergence of fundamental algorithms for feasibility problems. SIAM J. Optim. **23**(4), 2397–2419 (2013)
14. Hesse, R., Luke, D.R., Neumann, P.: Alternating projections and Douglas–Rachford for sparse affine feasibility. IEEE Trans. Signal. Process. **62**(18), 4868–4881 (2014)
15. Ioffe, A.D.: Metric regularity and subdifferential calculus. Russian Math. Surv. **55**(3), 501–558 (2000)
16. Ioffe, A.D.: Regularity on a fixed set. SIAM J. Optim. **21**(4), 1345–1370 (2011)
17. Ioffe, A.D.: Nonlinear regularity models. Math. Program. **139**(1–2), 223–242 (2013)
18. Ioffe, A.D.: Metric regularity: a survey. Part I. Theory. J. Aust. Math. Soc. **101**(2), 188–243 (2016)
19. Khanh, Phan Q., Kruger, A.Y., Thao, Nguyen H.: An induction theorem and nonlinear regularity models. SIAM J. Optim. **25**(4), 2561–2588 (2015)
20. Klatte, D., Kummer, B.: Nonsmooth Equations in Optimization. Kluwer, Dordrecht (2002)
21. Klatte, D., Kummer, B.: Optimization methods and stability of inclusions in Banach spaces. Math. Program. **117**(1–2), 305–330 (2009)
22. Kruger, A.Y.: Stationarity and regularity of set systems. Pac. J. Optim. **1**(1), 101–126 (2005)
23. Kruger, A.Y.: About regularity of collections of sets. Set-Valued Anal. **14**, 187–206 (2006)
24. Kruger, A.Y.: About stationarity and regularity in variational analysis. Taiwan. J. Math. **13**(6A), 1737–1785 (2009)
25. Kruger, A.Y.: Error bounds and metric subregularity. Optimization **64**(1), 49–79 (2015)
26. Kruger, A.Y., Luke, D.R., Thao, Nguyen H.: Set regularities and feasibility problems. Math. Program. B. https://doi.org/10.1007/s10107-016-1039-x
27. Kruger, A.Y., Luke, D.R., Thao, Nguyen H.: About subtransversality of collections of sets. Set-Valued Var. Anal. **25**(4), 701–729 (2017)
28. Kruger, A.Y., Thao, Nguyen H.: About uniform regularity of collections of sets. Serdica Math. J. **39**, 287–312 (2013)
29. Kruger, A.Y., Thao, Nguyen H.: About [q]-regularity properties of collections of sets. J. Math. Anal. Appl. **416**(2), 471–496 (2014)
30. Kruger, A.Y., Thao, Nguyen H.: Quantitative characterizations of regularity properties of collections of sets. J. Optim. Theory Appl. **164**, 41–67 (2015)

31. Kruger, A.Y., Thao, Nguyen H.: Regularity of collections of sets and convergence of inexact alternating projections. J. Convex Anal. **23**(3), 823–847 (2016)
32. Lewis, A.S., Luke, D.R., Malick, J.: Local linear convergence of alternating and averaged projections. Found. Comput. Math. **9**(4), 485–513 (2009)
33. Lewis, A.S., Malick, J.: Alternating projections on manifolds. Math. Oper. Res. **33**, 216–234 (2008)
34. Li, G., Pong, T.K.: Douglas–Rachford splitting for nonconvex feasibility problems. Math. Program. **159**(1), 371–401 (2016)
35. Luke, D.R.: Relaxed averaged alternating reflections for diffraction imaging. Inverse Problems **21**, 37–50 (2005)
36. Luke, D.R.: Finding best approximation pairs relative to a convex and a prox-regular set in Hilbert space. SIAM J. Optim. **19**(2), 714–739 (2008)
37. Luke, D.R.: ProxToolbox. http://num.math.uni-goettingen.de/proxtoolbox (2017). Accessed Aug 2017
38. Luke, D.R., Thao, Nguyen H., Tam, M.K.: Quantitative convergence analysis of iterated expansive, set-valued mappings. Math. Oper. Res. https://doi.org/10.1287/moor.2017.0898
39. Luke, D.R., Thao, Nguyen H., Teboulle, M.: Necessary conditions for linear convergence of Picard iterations and application to alternating projections. https://arxiv.org/pdf/1704.08926.pdf (2017)
40. Mordukhovich, B.S.: Variational Analysis and Generalized Differentiation. I: Basic Theory. Springer, Berlin (2006)
41. Moreau, J.-J.: Fonctions convexes duales et points proximaux dans un espace Hilbertien. Comptes Rendus de l'Académie des Sciences de Paris **255**, 2897–2899 (1962)
42. Noll, D., Rondepierre, A.: On local convergence of the method of alternating projections. Found. Comput. Math. **16**(2), 425–455 (2016)
43. Patrinos, P., Stella, L., Bemporad, A.: Douglas-Rachford splitting: Complexity estimates and accelerated variants. In: 53rd IEEE Conference on Decision and Control, pp. 4234–4239 (2014)
44. Penot, J.-P.: Calculus Without Derivatives. Springer, New York (2013)
45. Phan, H.M.: Linear convergence of the Douglas–Rachford method for two closed sets. Optimization **65**, 369–385 (2016)
46. Rockafellar, R.T., Wets, R.J.-B.: Variational Analysis. Grundlehren Math. Wiss. Springer, Berlin (1998)