Convergence of the Preconditioned Proximal Point Method and Douglas–Rachford Splitting in the Absence of Monotonicity

Brecht Evens   Pieter Pas   Puya Latafat   Panagiotis Patrinos

Abstract

The proximal point algorithm (PPA) is the most widely recognized method for solving inclusion problems and serves as the foundation for many numerical algorithms. Despite this popularity, its convergence results have been largely limited to the monotone setting. In this work, we study the convergence of (relaxed) preconditioned PPA for a class of nonmonotone problems that satisfy an oblique weak Minty condition. Additionally, we study the (relaxed) Douglas-Rachford splitting (DRS) method in the nonmonotone setting by establishing a connection between DRS and the preconditioned PPA with a positive semidefinite preconditioner. To better characterize the class of problems covered by our analysis, we introduce the class of semimonotone operators, offering a natural extension to (hypo)monotone and co(hypo)monotone operators, and describe some of their properties. Sufficient conditions for global convergence of DRS involving the sum of two semimonotone operators are provided. Notably, it is shown that DRS converges even when the sum of the involved operators (or of their inverses) is nonmonotone. Various example problems are provided, demonstrating the tightness of our convergence results and highlighting the wide range of applications our theory is able to cover.

Keywords. Convex/nonconvex optimization · monotone/nonmonotone variational inequalities · inclusion problems · preconditioned proximal point algorithm · Douglas–Rachford splitting · semimonotone operators
AMS subject classifications. 47H04 · 49J52 · 49J53 · 65K05 · 65K15 · 90C26.

1   Introduction

The preconditioned proximal point algorithm (PPPA) is one of the most fundamental methods for solving inclusion problems arising in optimization and variational analysis. It aims to find a solution to the inclusion problem

$$\text{find} \quad z \in \mathbb{R}^n \quad \text{such that} \quad 0 \in Tz, \tag{G-I}$$

where $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is a set-valued operator which is usually assumed to be monotone. The iterates of the (relaxed) preconditioned proximal point algorithm (PPPA) can be expressed as follows:

$$\begin{cases} \bar{z}^k \in (P + T)^{-1} P z^k \\ z^{k+1} = z^k + \lambda_k(\bar{z}^k - z^k). \end{cases} \tag{PPPA}$$

Here, $P \in \mathbb{R}^{n \times n}$ is a symmetric positive semidefinite matrix and $(\lambda_k)_{k \in \mathbb{N}}$ is a sequence of strictly positive relaxation parameters. The classic proximal point algorithm (PPA) [56] corresponds to the setting where the preconditioner P takes the form of $1/\gamma I$, where γ is a positive stepsize parameter, and the relaxation parameter λ_k is equal to one.

Several methods are known to be instances of PPPA, each with a specific choice of preconditioner aimed at exploiting structure present in the problem. These include well-known techniques such as Douglas-Rachford splitting (DRS) [31, Thm. 6], the alternating direction method of multipliers (ADMM) [28, §8], the (proximal) augmented Lagrangian method (ALM) [73, §4 & 5], the primal-dual hybrid gradient (PDHG) method [42,

Lem. 2.2], [32] (also known as Chambolle-Pock (CP) [18]), iterative refinement [64, §4.1.2], the method of partial inverses [31, §5], progressive hedging [76, Thm. 1], and progressive decoupling [75, Thm. 4].

The convergence analysis of the methods mentioned above typically relies upon an underlying monotonicity/convexity assumption [73, 74, 75]. See also [82, 55, 30, 17, 34] for related works and generalizations. Departing from this classical setting, we aim to establish convergence of relaxed PPPA for a class of nonmonotone operators that are defined through an *oblique weak Minty condition* (see Definition 2.1). This definition involves a symmetric matrix V that is not required to be positive or negative definite and controls the level of (non)monotonicity. When V is equal to the zero matrix, this condition reduces to the Minty variational inequality (MVI) [58, 35], sometimes also referred to as variational coherence, satisfied for instance by all pseudoconvex and star-convex functions [87]. When V takes the form of ρI for negative ρ, one obtains the so-called weak MVI that has been employed in the context of the extragradient method and forward-backward-forward splitting for solving nonconvex-nonconcave min-max problems [24, 68, 69, 12, 1, 2]. Weak MVI is closely related to the notion of cohypomonotonicity [10, 50, 85], for which a convergence analysis of classic PPA was already performed in [67, 46, 19]. The key difference between the weak MVI and cohypomonotonicity is that the governing inequality is not considered between any two points in the graph of T, but instead only between any point in the graph of T and (a subset of) the zeros of T (see also Example 2.6).

In contrast to the traditional convergence analysis relying on firm nonexpansiveness of the resolvent mapping, our convergence analysis leverages a projective interpretation that was introduced in the monotone case in [81, 80, 47]. We show that the same idea can be applied in the nonmonotone setting for the preconditioned proximal point algorithm provided that a different halfspace is considered. More specifically, each iteration of PPPA is interpreted as a projection onto a certain halfspace, separating the current iterate from a subset of solutions that satisfy the oblique weak Minty assumption. A second key aspect of our convergence analysis is to allow for positive semidefinite preconditioners (as opposed to positive definite ones), which is crucial for establishing a connection with splitting techniques such as (relaxed) Douglas-Rachford splitting (see Section 3.1). We note that in the monotone setting convergence of PPPA with semidefinite preconditioners has been studied in [49, Thm. 3.4], [13, §2.1]. In the nonmonotone setting, the aforementioned halfspace will be constructed using so-called *shadow sequences*, obtained by projecting the sequences $(z^k)_{k \in \mathbb{N}}$ and $(\bar{z}^k)_{k \in \mathbb{N}}$ onto the range of the preconditioner P.

In the second part of this paper, we consider the class of Douglas/Peaceman-Rachford splitting methods and shift our attention towards structured inclusion problems of the form

$$\text{find} \quad x \in \mathbb{R}^n \quad \text{such that} \quad 0 \in T_{\mathrm{P}} x := Ax + Bx, \tag{P-I}$$

where $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ are two (possibly nonmonotone) operators. Given stepsize $\gamma > 0$, strictly positive relaxation parameters $\lambda_k > 0$, and a certain initial guess $s^0 \in \mathbb{R}^n$, the (relaxed) Douglas-Rachford splitting method consists of the following iterates:

$$\begin{cases} u^k & \in J_{\gamma A}(s^k) \\ v^k & \in J_{\gamma B}(2u^k - s^k) \\ s^{k+1} = s^k + \lambda_k(v^k - u^k). \end{cases} \tag{DRS}$$

For $\lambda_k = 1$, this method corresponds to classic Douglas-Rachford splitting [26], while for $\lambda_k = 2$, it reduces to Peaceman-Rachford splitting (PRS) [66]. Initially introduced to solve systems of linear equations emerging in heat conduction problems, these methods were later extended in [54] for finding zeros of the sum of two maximally monotone operators, and convergence of the sequence $(u^k)_{k \in \mathbb{N}}$ to a solution was shown [54, 83].

In the optimization setting, global convergence of classic DRS (corresponding to $\lambda_k = 1$) in the nonconvex setting was obtained in [53] assuming that one function is Lipschitz differentiable. Similarly, PRS (corresponding to $\lambda_k = 2$) was considered in [52] under the additional requirement that the smooth function is strongly convex. These results were then unified in [84] and tight stepsize ranges for (relaxed) DRS in the nonconvex setting were provided. When restricting to their setting (one function being Lipschitz differentiable), we will show that when the nonsmooth term has a so-called *semimonotone* subdifferential, our theory allows for a larger stepsize range compared to [84] (see Remark 5.3*(iii)*).

A convergence analysis of DRS for minimizing the sum of an α-convex and a β-convex function has been provided in [41] under the assumption that $\alpha + \beta > 0$, i.e. in the convex case. Similar results were obtained for DRS applied to inclusion problems involving the sum of an α- and a β-monotone operator [21, 71] and the sum of an α- and a β-comonotone operator [8], both under the assumption that $\alpha + \beta > 0$. In the former work, it

is obvious that the so-called primal inclusion (P-I) is monotone, while in the latter, the so-called dual inclusion (D-I) is monotone due to [10, Lem. 2.6]. Therefore, the achieved results in this research direction have been limited to the case where either the primal or the dual problem is monotone.

1.1 Contributions

Our main contributions can be summarized as follows:

1. We establish convergence of PPPA for a class of nonmonotone operators that have oblique weak Minty solutions at one (or more) of its zeros (see Definition 2.1 and Theorem 2.4). Our convergence result hinges upon mere positive *semidefiniteness* of the preconditioning matrix (see Table 1), and tightness of our stepsize range is demonstrated in Example 2.6. Additionally, sublinear last-iterate convergence rates are obtained in Theorem 2.11 under a more stringent oblique comonotonicity assumption.

2. Leveraging a primal-dual connection between DRS and PPPA, convergence results of DRS are deduced whenever the associated primal-dual operator has oblique weak Minty solutions, as well as in the more restrictive, comonotone setting (see Section 3). Additionally, we demonstrate the tightness of this result in Example 6.1*(ii)*.

3. Despite the underlying nonmonotonicity, (local) linear convergence of PPPA is established in Theorem 2.13 under an additional metric subregularity assumption. In turn, this result leads to sufficient conditions for (local) linear convergence of DRS for (possibly) nonmonotone piecewise polyhedral mappings (see Theorem 3.6).

4. We introduce and develop calculus rules for the class of (μ, ρ)-semimonotone operators, which can be viewed as a natural extension of (hypo)monotone and co(hypo)monotone operators (see Remark 4.2). In addition, various examples of well-known (nonconvex) function classes with semimonotone subdifferentials are provided.

5. We provide sufficient conditions for the convergence of DRS, based on the semimonotonicity of the operators A and B (see Theorem 5.2). The stepsize region obtained through our analysis encompasses and extends existing results on relaxed Douglas/Peaceman-Rachford splitting methods such as [21, 8], see Remark 5.3. Up to the knowledge of the authors, this is the first work that is able to cover the convergence of DRS in the case where neither the primal, nor the dual, nor the primal-dual inclusion is monotone. We provide several example problems that demonstrate the applicability of our theory beyond the standard monotone setting.

1.2 Organization

The paper is structured as follows. Section 1.3 introduces some notation and recalls some standard definitions. In Section 2 the concept of oblique weak Minty solutions (cf. Definition 2.1) is formally introduced, and the convergence of PPPA is studied for a class of operators that admit at least one such solution. Section 3 establishes a primal-dual connection between PPPA and DRS, which leads to convergence results for DRS under an oblique weak Minty assumption on the associated primal-dual operator. In Section 4 we introduce the class of semimonotone operators, for which we provide several calculus rules and examples. Sufficient conditions for the convergence of DRS, based on the semimonotonicity of the underlying operators are provided in Section 5. In Section 6 various example problems covered by our theory for DRS are presented. Finally, Section 7 concludes the paper. For the sake of readability, several proofs and auxiliary results are deferred to the Appendix.

1.3 Notation

The set of natural numbers including zero is denoted by $\mathbb{N} := \{0, 1, \ldots\}$. The set of real and extended-real numbers are denoted by $\mathbb{R} := (-\infty, \infty)$ and $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$, while the positive and strictly positive reals are $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{R}_{++} := (0, \infty)$. We denote the positive part of a real number by $[\cdot]_+ := \max\{0, \cdot\}$ and the negative part by $[\cdot]_- := \min\{0, \cdot\}$. With id we indicate the identity function $x \mapsto x$ defined on a suitable space. We denote by $\mathbb{R}^n$ the standard n-dimensional Euclidean space with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$. For a vector $w = (w_1, \ldots, w_N) \in \mathbb{R}^n$, $w_i \in \mathbb{R}^{n_i}$ is used to denote its i-th (block) coordinate. The identity matrix is

denoted by $I_n \in \mathbb{R}^{n \times n}$; we write I when no ambiguity occurs. Given a matrix $P \in \mathbb{R}^{n \times n}$, we denote the range of A by $\mathcal{R}(A)$ and the kernel of A by $\mathcal{N}(A)$. The sets of symmetric, symmetric positive semidefinite and symmetric positive definite n-by-n matrices are denoted by $\mathbb{S}^n$, $\mathbb{S}^n_+$ and $\mathbb{S}^n_{++}$, respectively. We also write $P \geq 0$ and $P > 0$ for $P \in \mathbb{S}^n_+$ and $P \in \mathbb{S}^n_{++}$, respectively. We say that matrices $P, Q \in \mathbb{R}^{n \times n}$ are similar if there exists a nonsingular matrix $X \in \mathbb{R}^{n \times n}$ such that $P = X^{-1}QX$. For any matrix $P \in \mathbb{R}^{n \times n}$ with real eigenvalues, we denote its smallest eigenvalue by $\lambda_{\min}(P)$. Given a symmetric matrix $P \in \mathbb{S}^n$ we define the scalar product $\langle x, y \rangle_P = \langle x, Py \rangle$ and the quadratic function $q_P(x) := \langle x, x \rangle_P$. If $P \in \mathbb{S}^n_{++}$ then its associated induced norm is defined as $\|x\|_P = \sqrt{q_P(x)}$. We denote the Kronecker product between two matrices of arbitrary size by $\otimes$.

We use the notation $(z^k)_{k \in I}$ to denote a sequence with indices in the set $I \subseteq \mathbb{N}$. When dealing with scalar sequences we use the subscript notation $(\gamma_k)_{k \in I}$. We say that $(z^k)_{k \in \mathbb{N}}$ converges (sublinearly) with a big-$\mathcal{O}(g(k))$ rate if there exists an index $k_0 \in \mathbb{N}$ and a positive scalar c such that $|z^k| \leq cg(k)$ for all $k \geq k_0$ and we say that $(z^k)_{k \in \mathbb{N}}$ converges (sublinearly) with a little-$o(g(k))$ rate if $z^k/g(k) \to 0$. We say that $(z^k)_{k \in \mathbb{N}}$ converges to a point $z^\star$ (at least) Q-linearly with Q-factor given by $c \in (0, 1)$ if there exists an index $k_0 \in \mathbb{N}$ such that $\|z^{k+1} - z^\star\| \leq c\|z^k - z^\star\|$ for all $k \geq k_0$. We say that $(z^k)_{k \in \mathbb{N}}$ converges to a point $z^\star$ (at least) R-linearly if there exists a sequence of nonnegative scalars $(v^k)_{k \in \mathbb{N}}$ such that $\|z^k - z^\star\| \leq v^k$ and $(v^k)_{k \in \mathbb{N}}$ converges Q-linearly to zero.

An operator or set-valued mapping $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^d$ maps each point $x \in \mathbb{R}^n$ to a subset $A(x)$ of $\mathbb{R}^d$. We will use the notation $A(x)$ and Ax interchangeably. We denote the domain of A by $\operatorname{dom} A := \{x \in \mathbb{R}^n \mid Ax \neq \emptyset\}$, its graph by $\operatorname{gph} A := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^d \mid y \in Ax\}$, and the set of its zeros by $\operatorname{zer} A := \{x \in \mathbb{R}^n \mid 0 \in Ax\}$. The inverse of A is defined through its graph, i.e., $\operatorname{gph} A^{-1} := \{(y, x) \mid (x, y) \in \operatorname{gph} A\}$, and we denote the range of A by $\mathcal{R}(A) := \operatorname{dom} A^{-1}$. The *resolvent* of A is defined as $J_A := (\operatorname{id} + A)^{-1}$. The composition of two operators A and B is denoted by $A \circ B$.

Definition 1.1 (parallel sum of operators)**.** *The parallel sum between operators $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined as $A \square B := (A^{-1} + B^{-1})^{-1}$.*

Definition 1.2 (parallel sum of extended-real numbers)**.** *Let $a, b \in \overline{\mathbb{R}}$. We say that a and b are parallel summable if either $a = b = 0$ or $a + b \neq 0$ and their parallel sum is defined as*

$$a \square b := \begin{cases} 0, & \text{if } a = b = 0, \\ \frac{ab}{a+b}, & \text{otherwise,} \end{cases}$$

where we use the convention that $a \square \infty = a$.

Definition 1.3 ((co)monotonicity)**.** *An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be μ-monotone for some $\mu \in \mathbb{R}$ if*

$$\langle x - x', y - y' \rangle \geq \mu\|x - x'\|^2, \qquad \text{for all } (x, y), (x', y') \in \operatorname{gph} A,$$

and it is said to be ρ-comonotone for some $\rho \in \mathbb{R}$ if

$$\langle x - x', y - y' \rangle \geq \rho\|y - y'\|^2, \qquad \text{for all } (x, y), (x', y') \in \operatorname{gph} A.$$

The operator A is said to be maximally (co-)monotone if its graph is not strictly contained in the graph of another (co-)monotone operator.

We say that A is *outer semicontinuous (osc)* at $\bar{x} \in \operatorname{dom} A$ if

$$\limsup_{x \to \bar{x}} Ax := \{y \mid \exists x^k \to \bar{x}, \exists y^k \to y \text{ with } y^k \in Ax^k\} \subseteq A\bar{x}. \tag{1.1}$$

Outer semicontinuity of A everywhere is equivalent to its graph being a closed subset of $\mathbb{R}^n \times \mathbb{R}^d$.

The domain of an extended real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is the set $\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. We say that f is proper if $\operatorname{dom} f \neq \emptyset$ and that f is lower semicontinuous (lsc) if the epigraph $\operatorname{epi} f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}$ is a closed subset of $\mathbb{R}^{n+1}$. We denote the limiting subdifferential of f by ∂f. We say that f is ℓ-smooth to indicate that f is continuously differentiable and ∇f is Lipschitz continuous with modulus ℓ.

The indicator function of a set $E \subseteq \mathbb{R}^n$ is denoted by δ_E, namely $\delta_E(x) = 0$ if $x \in E$ and ∞ otherwise. We denote the normal cone of E by N_E. The projection onto and the distance from E with respect to $\|\cdot\|_Q$, $Q \in \mathbb{S}^n_{++}$, are denoted by

$$\Pi^Q_E(x) := \arg\min_{z \in E}\{\|z - x\|_Q\}, \quad \operatorname{dist}_Q(x, E) := \inf_{z \in E}\{\|z - x\|_Q\},$$

respectively. The absence of super/subscript Q implies the same definitions with respect to the canonical norm.

2 The preconditioned proximal point method

In this section, the convergence of PPPA with positive semidefinite preconditioning is studied for a class of nonmonotone operators, defined through an oblique weak Minty assumption.

We begin by introducing this class of nonmonotone operators and detailing our underlying assumptions. Then, leveraging a projective interpretation of the proximal point algorithm, we provide the corresponding convergence results of PPPA. Subsequently, we demonstrate the tightness of our convergence results through a simple example. Finally, we establish (local) linear convergence under an additional metric subregularity assumption.

2.1 Convergence analysis of PPPA under oblique weak Minty

Consider the following class of operators with oblique weak Minty solutions, which generalizes the class of operators with weak Minty solutions that has been employed in the context of the extragradient method and forward-backward-forward splitting [24, 69, 12].

Definition 2.1 (*V*-oblique weak Minty solutions). *An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to have V-oblique weak Minty solutions at (a nonempty set) $\mathcal{S}^\star \subseteq \operatorname{zer} T$ for some symmetric matrix $V \in \mathbb{S}^n$ if*

$$\langle v, z - z^\star \rangle \ge \mathsf{q}_V(v), \qquad \text{for all } z^\star \in \mathcal{S}^\star, (z, v) \in \operatorname{gph} T, \tag{2.1}$$

where the quadratic form $\mathsf{q}_V(v) := \langle v, Vv \rangle$. Whenever $V = \rho\mathrm{I}$ for some $\rho \in \mathbb{R}$, we refer to them as ρ-weak Minty solutions.

The generalization of the weak MVI to include a general symmetric matrix V instead of a scalar value ρ permits a more detailed characterization of problem classes, which in turn will lead to tight results for the convergence of PPPA and DRS (cf. Section 3).

The above definition can be further relaxed by requiring (2.1) to hold instead on $(z, v) \in \operatorname{gph} T \cap (\mathcal{R}((P + T)^{-1}P) \times \mathcal{R}(P))$, where P is the preconditioner from PPPA. Under this weaker condition, the results from Theorems 2.4 and 2.13 are still satisfied since (2.1) is invoked only at points in this restricted set (see (2.9)). We refrain from using this relaxation given that it couples the problem class with the preconditioner and does not appear to lead to tighter sufficient conditions in terms of semimonotonicity (cf. Section 5).

As previously mentioned in the introduction, a key aspect of our forthcoming convergence analysis for PPPA is to consider positive semidefinite preconditioners, largely inspired by the convergence analysis of [49, Thm. 3.4] and [20, Thm. 3.3]. Unlike the aforementioned works, we do not assume that T is the sum of a maximally monotone and a linear skew-symmetric operator. Instead, we require that it admits at least a single oblique weak Minty solution. More specifically, we work under the following assumptions.

Assumption I. *The operator T in (G-I) and the preconditioner P in (PPPA) satisfy the following properties.*

A1 $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ *is outer semicontinuous.*

A2 *The preconditioned resolvent $(P + T)^{-1}P$ has full domain.*

A3 *There exists a nonempty set $\mathcal{S}^\star \subseteq \operatorname{zer} T$ and a symmetric, possibly indefinite matrix $V \in \mathbb{S}^n$ such that T has V-oblique weak Minty solutions at $\mathcal{S}^\star$ for V.*

A4 $P \in \mathbb{S}^n_+$ *is a symmetric positive semidefinite matrix such that*

$$\eta_{\min} := 1 + \lambda_{\min}(U^\top VPU) > 0, \tag{2.2}$$

where U is any orthonormal basis for the range of P.

Note that the preconditioned resolvent $(P+T)^{-1}\circ P$ can be seen as a particular instance of the warped/nonlinear resolvent considered in the monotone setting in [15, 36].

Remark 2.2. Assumption I.A4 provides a restriction on the preconditioner P, depending on the oblique weak Minty matrix V. To better understand this condition, consider the matrix $X := U^\top PU$, which is positive definite

5
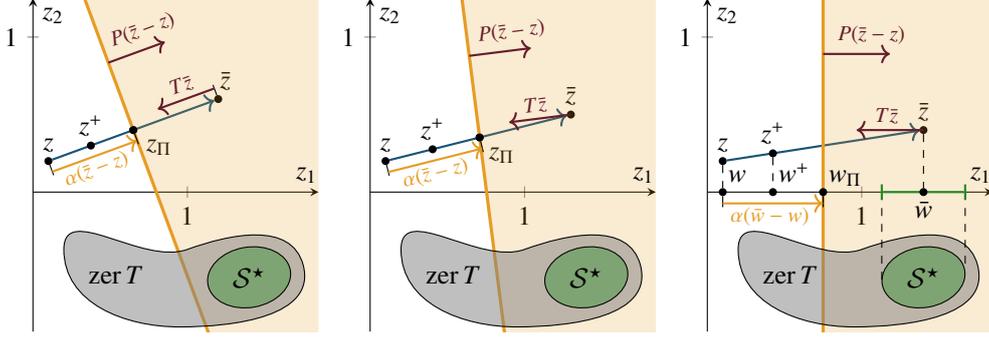
Figure 1: *Visualization of the involved variables in Lemma 2.3 for different types of preconditioners, where $V = -\frac{3}{2}I$, α is as in (2.7) and $\lambda = \alpha/2$. Here, we use the shorthand notation $w_\Pi = \Pi^Q_{\mathcal{D}_{w,\bar{w}}}(w)$ and $z_\Pi = \Pi^P_{\mathcal{D}_{z,\bar{z}}}(z)$. In these figures, the vector $\bar{z}$ satisfies the update rule (PPPA) (see (2.8)). Note that the precise location of the halfspace is governed by the oblique weak Minty matrix V, due to the dependence of α on V. For example, if $V = 0$, then $\alpha = 1$ and the corresponding hyperplane would pass through $\bar{w}$. (left and middle) Positive definite preconditioner, such that $\Pi_{\mathcal{R}(P)} = I$, $P = Q$ and the variables $z, \bar{z}, z^+, z_\Pi$ are equal to the variables $w, \bar{w}, w^+, w_\Pi$. (left) Multiple of identity preconditioner $P = \frac{1}{3}I$. (middle) $P = \mathrm{diag}(1/3, 1/6)$. (right) Positive semidefinite preconditioner $P = \mathrm{diag}(1/3, 0)$.*

by construction (see Lemma A.1*(iv)*). By similarity transformation using the similarity matrix $X^{1/2}$ [44, Cor. 1.3.4], it follows that $U^\top VPU$ is similar to the symmetric matrix $X^{1/2}U^\top VUX^{1/2}$, i.e., that

$$\eta_{\min} = 1 + \lambda_{\min}(X^{1/2}U^\top VUX^{1/2}). \tag{2.3}$$

As a result, Assumption I.A4 holds if and only if

$$(U^\top PU)^{-1} + U^\top VU \succ 0. \tag{2.4}$$

This condition is vacuously satisfied in the case where $U^\top VU \succeq 0$, which includes the WMI setting. The preconditioner P needs to be selected properly only if $U^\top VU$ is not positive semidefinite. For instance, when $P = 1/\gamma I$ where $\gamma > 0$, Assumption I.A4 reduces to $\gamma > [-\lambda_{\min}(V)]_+$. Moreover, when $V = \rho I$, it matches the stepsize condition $\gamma > [-\rho]_+$ for cohypomonotone operators (see e.g. [19, Thm. 3.1] and Table 1 for more details). □

Our convergence analysis relies on a projective interpretation of the preconditioned proximal point algorithm in the monotone setting that dates back to [81, 80, 47]. In particular, it was shown that when the preconditioner is positive definite, each iteration of PPPA can be interpreted as a projection onto a certain halfspace, constructed using the sequences $(z^k)_{k\in\mathbb{N}}$ and $(\bar{z}^k)_{k\in\mathbb{N}}$. In particular, under the assumption that T is maximally monotone and $P = \gamma^{-1}I$, [80] considers the halfspace

$$\mathcal{D}_{z,\bar{z}} := \left\{ r \in \mathbb{R}^n \mid \langle \gamma^{-1}(z - \bar{z}), \bar{z} - r \rangle \geq 0 \right\}. \tag{2.5}$$

We will show that this idea can be extended to our more general setting provided that a different halfspace is considered. More specifically, each iteration of PPPA is interpreted as a projection onto a halfspace separating the current iterate from a subset of solutions that satisfy the oblique weak Minty assumption. This halfspace will be constructed using the so-called shadow sequences $(w^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ and $(\bar{w}^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} \bar{z}^k)_{k\in\mathbb{N}}$ to deal with the fact that in our work the preconditioner is only positive *semidefinite*.

We proceed to present this separating halfspace and its main properties in the following lemma, for which a supporting visualization is provided in Figure 1. Note that the introduction of the matrix $Q := P + \Pi_{\mathcal{N}(P)}$ is inspired by the convergence analysis performed in [49, Thm. 3.4] and [20, Thm. 3.3].

Lemma 2.3 (halfspace interpretation). *Suppose that Assumption I holds. Consider the points $z \in \mathbb{R}^n$, $\bar{z} \in (P + T)^{-1}Pz$ and $z^+ = z + \lambda(\bar{z} - z)$ satisfying update rule (PPPA), and let $w := \Pi_{\mathcal{R}(P)} z$, $\bar{w} := \Pi_{\mathcal{R}(P)} \bar{z}$ and $w^+ := \Pi_{\mathcal{R}(P)} z^+$. Define the matrix $Q := P + \Pi_{\mathcal{N}(P)}$ and the halfspace*

$$\mathcal{D}_{w,\bar{w}} := \{ r \in \mathbb{R}^n \mid \langle w - \bar{w}, \bar{w} - r \rangle_Q \geq q_V(Q(w - \bar{w})) \}. \tag{2.6}$$

Then, the following hold.

6

(i) The set $\Pi_{\mathcal{R}(P)}\,\mathcal{S}^\star$ is a subset of $\mathscr{D}_{w,\bar{w}}$.

(ii) If $w \in \mathscr{D}_{w,\bar{w}}$ then $P(z - \bar{z}) = 0$, and in particular $\bar{z} \in \operatorname{zer} T$.

(iii) If $\bar{z} \notin \operatorname{zer} T$, then $\mathrm{q}_P(z - \bar{z}) \neq 0$ and

$$\alpha := 1 + \frac{\mathrm{q}_V(P(z - \bar{z}))}{\mathrm{q}_P(z - \bar{z})} \geq \eta_{\min} > 0. \tag{2.7}$$

(iv) $w^+ = (1 - \lambda/\alpha)w + \lambda/\alpha\,\Pi^Q_{\mathscr{D}_{w,\bar{w}}}(w)$ for any $\lambda \in \mathbb{R}$.

Proof.

♠ 2.3*(i)*: By update rule (PPPA) it holds that

$$P(z - \bar{z}) \in T\bar{z}. \tag{2.8}$$

Therefore, owing to the V-oblique weak Minty assumption (Assumption I.A3), for all $z^\star \in \mathcal{S}^\star$

$$\langle P(z - \bar{z}),\, \bar{z} - z^\star \rangle \geq \mathrm{q}_V(P(z - \bar{z})). \tag{2.9}$$

As a result, for all $w^\star := \Pi_{\mathcal{R}(P)}\, z^\star \in \Pi_{\mathcal{R}(P)}\,\mathcal{S}^\star \subseteq \Pi_{\mathcal{R}(P)} \operatorname{zer} T$

$$\langle Q(w - \bar{w}),\, \bar{w} - w^\star \rangle = \langle P(z - \bar{z}),\, \bar{z} - z^\star \rangle \overset{(2.9)}{\geq} \mathrm{q}_V(P(z - \bar{z})) = \mathrm{q}_V(Q(w - \bar{w})),$$

where we used Lemma A.1*(ii)* to relate P and Q in the first and final equality. It is now evident that by construction, $\Pi_{\mathcal{R}(P)}\,\mathcal{S}^\star \subseteq \mathscr{D}_{w,\bar{w}}$.

♠ 2.3*(ii)*: Suppose that $w \in \mathscr{D}_{w,\bar{w}}$ and define the shorthand notation $\tilde{w} := w - \bar{w}$, $\tilde{z} := z - \bar{z}$. Owing to Lemmas A.1*(i)* and A.1*(ii)*, it holds that

$$P = Q\,\Pi_{\mathcal{R}(P)} = \Pi_{\mathcal{R}(P)}\, Q\, \Pi_{\mathcal{R}(P)} = UU^\top QUU^\top = UXU^\top. \tag{2.10}$$

Consequently, by definition of $\mathscr{D}_{w,\bar{w}}$, it holds that

$$\begin{aligned}
0 \geq \|\tilde{w}\|^2_Q + \mathrm{q}_V(Q\tilde{w}) &= \langle Q\,\Pi_{\mathcal{R}(P)}\,\tilde{w},\, (\mathrm{I} + VQ\,\Pi_{\mathcal{R}(P)})\tilde{w} \rangle \\
&= \langle UXU^\top\tilde{w},\, (\mathrm{I} + VUXU^\top)\tilde{w} \rangle = \langle X^{1/2}U^\top\tilde{w},\, X^{1/2}U^\top(\mathrm{I} + VUXU^\top)\tilde{w} \rangle \\
&= \langle X^{1/2}U^\top\tilde{w},\, (\mathrm{I} + X^{1/2}U^\top VUX^{1/2})X^{1/2}U^\top\tilde{w} \rangle. \tag{2.11}
\end{aligned}$$

Furthermore, due to Assumption I.A4 and Remark 2.2 it follows that $\mathrm{I} + X^{1/2}U^\top VUX^{1/2}$ is positive definite, which combined with (2.11) and $X \succ 0$ implies that $U^\top\tilde{w} = 0$. In turn, using Lemmas A.1*(i)* and A.1*(ii)* we have $P\bar{z} = P\,\Pi_{\mathcal{R}(P)}\,\tilde{w} = PUU^\top\tilde{w} = 0$, which by (2.8) implies that $\bar{z} \in \operatorname{zer} T$.

♠ 2.3*(iii)*: If $\bar{z} \notin \operatorname{zer} T$, then inclusion (2.8) implies that $P(z - \bar{z}) \neq 0$. Combined with $Q(w - \bar{w}) = P(z - \bar{z})$ (where we used Lemma A.1*(ii)*) and positive definiteness of Q (see Lemma A.1*(iii)*) it follows that $\mathrm{q}_P(z - \bar{z}) = \|w - \bar{w}\|^2_Q > 0$, ensuring that $\alpha = 1 + \frac{\mathrm{q}_V(Q\tilde{w})}{\|\tilde{w}\|^2_Q}$ is finite-valued. Consequently, using (2.11) and that $\|\tilde{w}\|^2_Q = \langle Q\,\Pi_{\mathcal{R}(P)}\,\tilde{w},\, \tilde{w} \rangle = \|X^{1/2}U^\top\tilde{w}\|^2$ by (2.10), α may be written as a Rayleigh quotient [44, Thm. 4.2.2] of $\mathrm{I} + X^{1/2}U^\top VUX^{1/2}$ and $X^{1/2}U^\top\tilde{w}$, i.e.,

$$\alpha = 1 + \frac{\mathrm{q}_V(Q\tilde{w})}{\|\tilde{w}\|^2_Q} = \frac{\langle X^{1/2}U^\top\tilde{w},\, (\mathrm{I} + X^{1/2}U^\top VUX^{1/2})X^{1/2}U^\top\tilde{w} \rangle}{\|X^{1/2}U^\top\tilde{w}\|^2} \geq \lambda_{\min}(\mathrm{I} + X^{1/2}U^\top VUX^{1/2}) \overset{(2.3)}{=} \eta_{\min}.$$

♠ 2.3*(iv)*: For any $u \notin \mathscr{D}_{w,\bar{w}}$, it holds that $\Pi^Q_{\mathscr{D}_{w,\bar{w}}}(u) = u + \frac{\langle w - \bar{w}, u - \bar{w} \rangle_Q + \mathrm{q}_V(Q(w - \bar{w}))}{\|w - \bar{w}\|^2_Q}(\bar{w} - w)$ [9, Ex. 29.20]. Therefore, $\Pi^Q_{\mathscr{D}_{w,\bar{w}}}(w) = w + \alpha(\bar{w} - w)$ and for any $\lambda \in \mathbb{R}$ it holds that

$$(1 - \lambda/\alpha)w + \lambda/\alpha\,\Pi^Q_{\mathscr{D}_{w,\bar{w}}}(w) = w + \lambda(\bar{w} - w) = \Pi_{\mathcal{R}(P)}\,(z + \lambda(\bar{z} - z)) = \Pi_{\mathcal{R}(P)}\, z^+ = w^+.$$

$\square$

Most notably, Lemma 2.3*(iv)* establishes that the update rule for the (shadow) sequence generated by PPPA can be interpreted as a relaxed projection onto the halfspace defined in (2.6), while Lemmas 2.3*(i)* and 2.3*(ii)* imply that if at any iteration the current iterate belongs to the halfspace containing the set of projected oblique weak Minty solutions $\Pi_{\mathcal{R}(P)} \mathcal{S}^\star$, this implies its optimality. Owing to the non-expansiveness of the projection operator, this leads to the following convergence result for the preconditioned proximal point algorithm.

Theorem 2.4 (convergence of PPPA). *Suppose that Assumption I holds, and consider a sequence* $(z^k, \bar{z}^k)_{k\in\mathbb{N}}$ *generated by PPPA starting from* $z^0 \in \mathbb{R}^n$ *with relaxation parameters* $\lambda_k \in (0, 2\alpha_k)$ *such that* $\liminf_{k\to\infty} \lambda_k(2\alpha_k - \lambda_k) > 0$, *where*

$$\alpha_k = 1 + \frac{q_V(P(z^k - \bar{z}^k))}{q_P(z^k - \bar{z}^k)}. \tag{2.12}$$

Then, either a point $\bar{z}^k \in \operatorname{zer} T$ *is reached in a finite number of iterations or the following hold for the sequence* $(z^k, \bar{z}^k)_{k\in\mathbb{N}}$.

  (i) *$\bar{v}^k := P(z^k - \bar{z}^k) \in T\bar{z}^k$ for all k and $(\bar{v}^k)_{k\in N}$ converges to zero.*

  (ii) *Every limit point (if any) of* $(\bar{z}^k)_{k\in\mathbb{N}}$ *belongs to* $\operatorname{zer} T$.

  (iii) *The shadow sequences* $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$, $(\Pi_{\mathcal{R}(P)} \bar{z}^k)_{k\in\mathbb{N}}$ *are bounded and their limit points belong to* $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$.

  (iv) *If $\lambda_k(2\alpha_k - \lambda_k) \geq \kappa > 0$ uniformly for all k then*

$$\min_{k=0,1,\dots,N} q_P(z^k - \bar{z}^k) \leq \frac{1}{(N+1)\kappa} q_P(z^0 - z^\star), \qquad \text{for all } z^\star \in \mathcal{S}^\star.$$

  (v) *If in Assumption I.A3 $\Pi_{\mathcal{R}(P)} \mathcal{S}^\star = \Pi_{\mathcal{R}(P)} \operatorname{zer} T$, then $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ converges to some element of $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$. If additionally $(P + T)^{-1} \circ P$ is (single-valued) continuous, then $(\bar{z}^k)_{k\in\mathbb{N}}$ converges to some $z^\star \in \operatorname{zer} T$. Finally, if λ_k is additionally uniformly bounded in the interval $(0, 2)$, then $(z^k)_{k\in\mathbb{N}}$ also converges to $z^\star \in \operatorname{zer} T$.*

Proof. Note that as long as $\bar{z}^k \notin \operatorname{zer} T$ it follows from Lemma 2.3*(iii)* that $\alpha_k \geq \eta_{\min} > 0$. Consider the shadow sequences $(w^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ and $(\bar{w}^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} \bar{z}^k)_{k\in\mathbb{N}}$ with corresponding update rule

$$w^{k+1} = w^k + \lambda_k(\bar{w}^k - w^k). \tag{2.13}$$

If $w^k \in \mathscr{D}_{w^k, \bar{w}^k}$ for some $k \in \mathbb{N}$ then by Lemma 2.3*(ii)* the algorithm has reached a point $\bar{z}^k \in \operatorname{zer} T$ in a finite number of iterations. Therefore, we will consider the case when $w^k \notin \mathscr{D}_{w^k, \bar{w}^k}$ in the remainder of the proof.

♠ 2.4*(i)*: From the update rule (PPPA), it directly follows that $\bar{v}^k := P(z^k - \bar{z}^k) \in T\bar{z}^k$ for all k. It remains to show that $(\bar{v}^k)_{k\in\mathbb{N}}$ converges to zero. By Lemma 2.3*(iv)*, update step (2.13) is equivalent to the relaxed projection $\tilde{\Pi}_k := (1 - \zeta_k)\mathrm{id} + \zeta_k \Pi^Q_{\mathscr{D}_{w^k, \bar{w}^k}}$, where $\zeta_k := \lambda_k/\alpha_k$. Owing to firm nonexpansiveness of $\Pi^Q_{\mathscr{D}_{w^k, \bar{w}^k}}$ [9, Prop. 4.16], the relaxed projection $\tilde{\Pi}_k$ is $\zeta_k/2$-averaged in the space with inner product $\langle \cdot, \cdot \rangle_Q$ [9, Cor. 4.41]. Note that for any $w^\star \in \Pi_{\mathcal{R}(P)} \mathcal{S}^\star$ it holds by Lemma 2.3*(i)* that $w^\star \in \mathscr{D}_{w^k, \bar{w}^k}$. Therefore, it follows from [9, Prop. 4.35(iii)] for any $w^\star \in \Pi_{\mathcal{R}(P)} \mathcal{S}^\star$ that

$$\begin{aligned}
\|w^{k+1} - w^\star\|_Q^2 = \|\tilde{\Pi}_k(w^k) - \tilde{\Pi}_k(w^\star)\|_Q^2 &\leq \|w^k - w^\star\|_Q^2 - \frac{1 - \zeta_k/2}{\zeta_k/2}\|w^{k+1} - w^k\|_Q^2 \\
&= \|w^k - w^\star\|_Q^2 - \frac{2\alpha_k - \lambda_k}{\lambda_k}\|\lambda_k(w^k - \bar{w}^k)\|_Q^2 \\
&= \|w^k - w^\star\|_Q^2 - \lambda_k(2\alpha_k - \lambda_k)\|w^k - \bar{w}^k\|_Q^2, \tag{2.14}
\end{aligned}$$

establishing that $(w^k)_{k\in\mathbb{N}}$ is Fejér monotone with respect to $\Pi_{\mathcal{R}(P)} \mathcal{S}^\star$ [9, Def. 5.1]. By telescoping (2.14), and since $\liminf_{k\to\infty} \lambda_k(2\alpha_k - \lambda_k) > 0$, it follows that $\sum_{k=0}^\infty \|w^k - \bar{w}^k\|_Q^2 < +\infty$, which implies that

$$\|w^k - \bar{w}^k\|_Q \to 0. \tag{2.15}$$

The convergence of $(\bar{v}^k)_{k\in N}$ to zero is established by noting that

$$\|\bar{v}^k\|^2 = \|P(w^k - \bar{w}^k)\|^2 \leq \|P^{1/2}\|^2 \|P^{1/2}(w^k - \bar{w}^k)\|^2 = \|P\| \|w^k - \bar{w}^k\|_Q^2. \tag{2.16}$$

Table 1: *Connection between Theorem 2.4 and existing convergence results for PPPA.*

		$P = \gamma^{-1}\mathrm{I}$		$P \geq 0$
	Conditions	Theorems	Conditions	Theorems
$V = 0$	$\gamma > 0, \lambda_k = 1$	[56][1], [74][1,2], [3, Thm. 3.2]	$\lambda_k \in (0,2)$	[49, Thm. 3.4][1], [13, Thm. 2.9][1]
	$\gamma > 0,$ $\lambda_k \in (0,2)$	[31, Thm. 3][1,2]		
$V = \rho\mathrm{I}$	$\gamma > [-2\rho]_+,$ $\lambda_k = 1$	[67, Thm. 9][1,2], [46, Thm. 1][1,2], [39, Thm. 3.1]	$1 + \lambda_{\min}(\rho U^\top P U) > 0,$ $\lambda_k \in (0, 2\alpha_k)$	Theorem 2.4
	$\gamma > [-\rho]_+,$ $\lambda_k \in \left(0, 2(1 + \tfrac{\rho}{\gamma})\right)$	[19, Thm. 3.1][1,2]		
$V \in \mathbb{S}$	$\gamma > [-\lambda_{\min}(V)]_+,$ $\lambda_k \in (0, 2\alpha_k)$	Theorem 2.4	$1 + \lambda_{\min}(U^\top V P U) > 0,$ $\lambda_k \in (0, 2\alpha_k)$	Theorem 2.4

[1]Considers maximal monotonicity instead of MVI or maximal ρ-comonotonicity instead of ρ-weak Minty.
[2]Considers the setting where the stepsize γ_k is variable and the resolvent computations are performed inexactly.

♠ *2.4(ii):* Since $(\bar{v}^k)_{k\in\mathbb{N}}$ converges to zero and $\bar{v}^k \in T\bar{z}^k$, it follows from outer semicontinuity of T (due to Assumption I.A1) that any limit point of $(\bar{z}^k)_{k\in\mathbb{N}}$ belongs to zer T.

♠ *2.4(iii):* It follows from (2.14) that $(\|w^k - w^\star\|_Q)_{k\in\mathbb{N}}$ converges, and in particular that $(w^k)_{k\in\mathbb{N}}$ is bounded. In turn, using (2.15) and the triangle inequality $\|\bar{w}^k\|_Q \leq \|\bar{w}^k - w^k\|_Q + \|w^k\|_Q$, it follows that $(\bar{w}^k)_{k\in\mathbb{N}}$ is bounded and thus that it has at least one limit point. Take a subsequence $(\bar{w}^k)_{k\in K}$ converging to some limit point w^∞. Since $(\bar{w}^k - w^k)_{k\in\mathbb{N}}$ converges to zero by (2.15), we have that $(w^k)_{k\in K}$ also converges to the same limit point w^∞. The claim is established by noting that $w^\infty \in \Pi_{\mathcal{R}(P)}\,\mathrm{zer}\,T$ owing to Theorem 2.4(ii).

♠ *2.4(iv):* By telescoping (2.14) and from the uniform bound $\forall k \in \mathbb{N} : \lambda_k(2\alpha_k - \lambda_k) \geq \kappa > 0$, it follows that

$$\min_{0 \leq k \leq N} \|w^k - \bar{w}^k\|_Q^2 \leq \tfrac{1}{(N+1)\kappa}\|w^0 - w^\star\|_Q^2. \tag{2.17}$$

The claimed upper bound thus follows by noting that $\|w^k - \bar{w}^k\|_Q^2 = \mathsf{q}_P(z^k - \bar{z}^k)$ and $\|w^0 - w^\star\|_Q^2 = \mathsf{q}_P(z^0 - z^\star)$.

♠ *2.4(v):* If $\Pi_{\mathcal{R}(P)}\mathcal{S}^\star = \Pi_{\mathcal{R}(P)}\,\mathrm{zer}\,T$, then the convergence of $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ to some element of $\Pi_{\mathcal{R}(P)}\,\mathrm{zer}\,T$ follows by [9, Thm. 5.5]. Let $G = (P + T)^{-1} \circ P$ be (single-valued) continuous, so that $\bar{z}^k = Gz^k = Gw^k$. Take a subsequence $(w^k)_{k\in K}$ converging to some limit point w^∞. Then, by continuity of G, the sequence $(\bar{z}^k)_{k\in\mathbb{N}}$ converges to $\bar{z}^\infty = Gw^\infty$, and by Theorem 2.4(ii) it holds that $\bar{z}^\infty \in \mathrm{zer}\,T$.

It remains to show that $(z^k)_{k\in\mathbb{N}}$ also converges to $\bar{z}^\infty$ if λ_k is uniformly bounded in the interval $(0,2)$. Note that

$$\|z^{k+1} - \bar{z}^\infty\| \leq \lambda_k\|\bar{z}^k - \bar{z}^\infty\| + |1 - \lambda_k|\|z^k - \bar{z}^\infty\|.$$

Consequently, by uniform boundedness of λ_k, i.e., $(\lambda_k)_{k\in\mathbb{N}} \subseteq [\epsilon, 2 - \epsilon]$ for some $\epsilon > 0$, and by [70, Lem. 3 of §2.2] it follows that $\lim_{k\to\infty}\|z^{k+1} - \bar{z}^\infty\| = 0$, i.e., that $(z^k)_{k\in\mathbb{N}}$ also converges to $\bar{z}^\infty \in \mathrm{zer}\,T$. □

Note that as long as the solution has not yet been reached, i.e., as long as $\bar{z}^k \notin \mathrm{zer}\,T$, then $\alpha_k \geq \eta_{\min} > 0$ by Lemma 2.3(iii). Therefore, all claims of Theorem 2.4 also hold for the (more restrictive) fixed stepsize range $\lambda_k \in (0, 2\eta_{\min})$.

Due to the generality of our underlying assumptions, Theorem 2.4 unifies and extends existing convergence results for PPPA in literature, as summarized in Table 1. One notable instance of Theorem 2.4 is when $V = \rho\mathrm{I}$ and $P = \gamma^{-1}\mathrm{I}$, in which case $\alpha_k = \eta_{\min} = 1 + \rho/\gamma$. For completeness, we present the simplified theorem statement for this setting below.

Corollary 2.5. *Suppose that T is outer semicontinuous, that there exists a nonempty set $\mathcal{S}^\star \subseteq \mathrm{zer}\,T$ such that T has ρ-weak Minty solutions at $\mathcal{S}^\star$ and that the resolvent $J_{\gamma T}$ has full domain given stepsize $\gamma > [-\rho]_+$. Consider a sequence $(z^k, \bar{z}^k)_{k\in\mathbb{N}}$ generated by PPPA starting from $z^0 \in \mathbb{R}^n$ with preconditioner $P = 1/\gamma\mathrm{I}$ and relaxation parameters $\lambda_k \in (0, 2(1 + \rho/\gamma))$ satisfying $\liminf_{k\to\infty} \lambda_k(2(1 + \rho/\gamma) - \lambda_k) > 0$. Then, all the claims from Theorem 2.4 hold.*

9
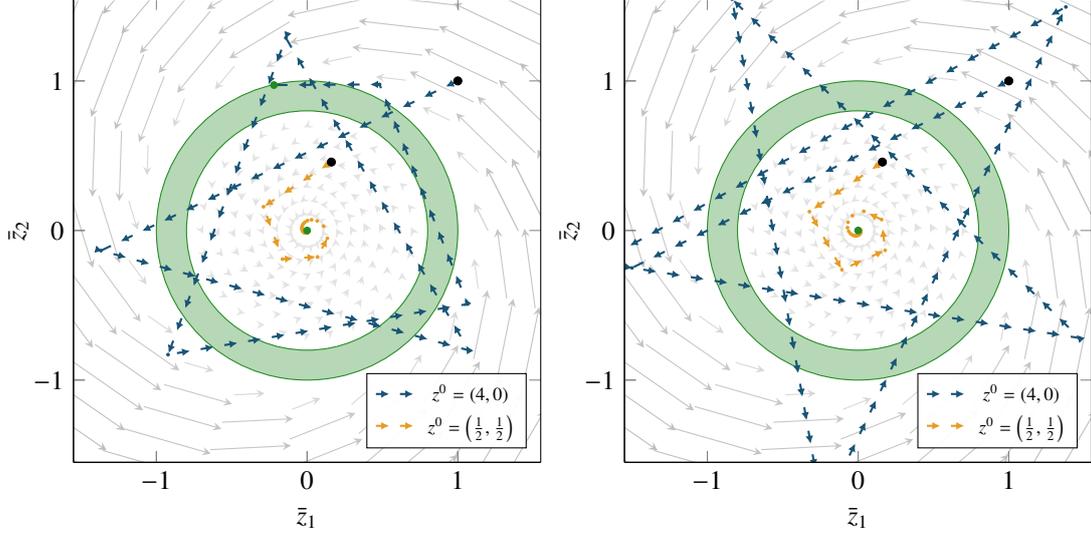
Figure 2: *Visualization of the sequences $(\bar{z}^k)_{k\in\mathbb{N}}$ for the relaxed proximal point algorithm with stepsize $\gamma = 1$ and fixed relaxation parameter λ applied to Example 2.6, where $a = 2$ and $b = 1$. The zeros of T are marked in green and its gradient flow $-T$ is indicated using gray arrows. (left) Sequences generated using $\lambda = 2.3 < \bar{\lambda}$, where $\bar{\lambda}$ is the upper bound on λ_k from (2.18). The orange sequence converges to the single (b/a^2+b^2)-weak Minty solution $(0,0)$ while the blue sequence converges to a zero of T which is not the weak Minty solution (cf. Theorem 2.4(ii)). (right) Sequences generated using $\lambda = 2.5 > \bar{\lambda}$. In this setting, it is no longer guaranteed by Theorem 2.4 that the sequence $(\bar{z}^k)_{k\in\mathbb{N}}$ converges, as can be seen from the (diverging) blue sequence.*

It is worth reiterating that all results from Theorem 2.4 except the final one apply even when only a subset of zer T are oblique weak Minty solutions. Most notably, Theorem 2.4(ii) states that the limit points of $(\bar{z}^k)_{k\in\mathbb{N}}$ are zeros of T, which do not have to be oblique weak Minty solutions. The significance of this result is highlighted by the second part of the following example, where this convergence behavior is observed in practice (see also Figure 2). The first part of this example demonstrates the tightness of our results by considering a simple linear inclusion related to saddle point problems. The details are deferred to Appendix C. Supplementary code for the numerical examples presented in this work can be found on GitHub.

Example 2.6 (toy example)**.** Consider the operator

$$T(z) := f(\|z\|)\begin{bmatrix} b & a \\ -a & b \end{bmatrix} z,$$
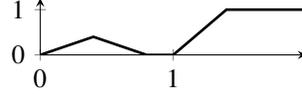
Figure 3: *Definition of $f : \mathbb{R}_+ \to [0, 1]$ in Example 2.6.*

where $f : \mathbb{R}_+ \to [0, 1]$ and $a, b \in \mathbb{R}$ are not simultaneously equal to zero.

(i) If $f \equiv 1$, then the following hold.

    1. T is (b/a^2+b^2)-comonotone, and thus T has a single (b/a^2+b^2)-weak Minty solution at $\mathcal{S}^\star = $ zer $T = (0, 0)$.

    2. By Theorem 2.4(v), the sequence $(\bar{z}^k)_{k\in\mathbb{N}}$ generated by the proximal point algorithm with fixed relaxation parameter λ converges to zer T if $\gamma > \frac{[-b]_+}{a^2+b^2}$ and λ lies in the interval

$$0 < \lambda < \bar{\lambda} := 2\left(1 + \frac{b}{\gamma(a^2+b^2)}\right). \tag{2.18}$$

    By examining the spectral radius of the algorithmic operator, it can be seen that this result is tight.

(ii) If f is defined as in Figure 3 and b is nonnegative, then the following hold.

    1. T has a single (b/a^2+b^2)-weak Minty solution at the point $\mathcal{S}^\star = (0, 0) \subset$ zer T while the other zeros of T are not ρ-weak Minty solutions for any $\rho \in \mathbb{R}$.

    2. By Theorem 2.4(ii), the limit points of the sequence $(\bar{z}^k)_{k\in\mathbb{N}}$ generated by the relaxed proximal point algorithm with fixed relaxation parameter λ belong to zer T (not necessarily to $\mathcal{S}^\star$) if $\gamma > 0$ and λ is selected according to (2.18). This result has been verified numerically in Figure 2.

10
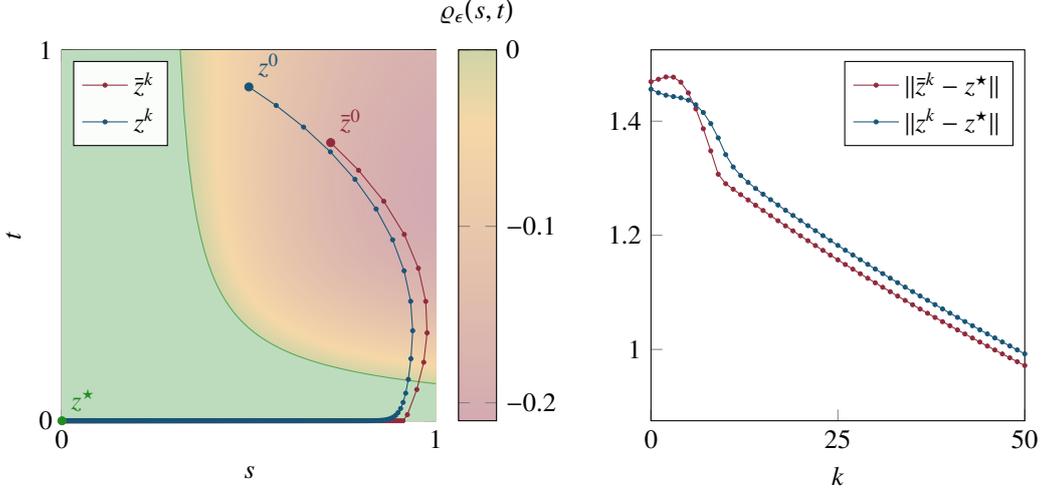
Figure 4: *(left) Visualization of the sequences $(z^k)_{k \in \mathbb{N}}$ and $(\bar{z}^k)_{k \in \mathbb{N}}$ for the relaxed proximal point algorithm, applied to the minimax problem from Example 2.7, with problem parameter $\epsilon = 1/10$, weak Minty constant $\rho = \frac{\epsilon-1}{4} = -9/40$, stepsize $\gamma = 1/4 > [-\rho]_+$ and relaxation parameter $\lambda = 1/3 \in (0, 2(1 + \rho/\gamma))$. These choices comply with Corollary 2.5. The domain $z \in \Delta_2 \times \Delta_2$ is parametrized by $z = (s, 1-s, t, 1-t)$. The heat map indicates the value of $\varrho_\epsilon(z) = \inf_{v \in N_C(z)} \langle F(z) + v, z - z^\star \rangle / \|F(z) + v\|^2$, which can be thought of as a local measure of nonmonotonicity of the operator (see also Appendix C). The region where $\varrho_\epsilon(z)$ is positive is marked in green. Even though the iterates pass through the nonmonotone region, both sequences converge to the Nash equilibrium. (right) Visualization of the distance between the sequences $(z^k)_{k \in \mathbb{N}}$ and $(\bar{z}^k)_{k \in \mathbb{N}}$ and the Nash equilbrium $z^\star$. Note that in the first iterations, when passing through the nonmonotone region, the sequence $\|\bar{z}^k - z^\star\|$ increases. On the other hand, the sequence $\|z^k - z^\star\|$ is monotonically decreasing (see also (2.14)).*

$\square$

As a second illustrative example, we will consider the application of our theory to the economic equilibrium model of Von Neumann [61]. For this example, it was recently observed in [22, Prop. 2] that for certain choices R and S, operator F is nonmonotone but satisfies the weak MVI. Here, we extend this result by showing that the Nash equilibrium of (2.19) is a weak Minty solution of the full operator $T := F + N_C$. The details are deferred to Appendix C.

Example 2.7 (Von Neumann's economic equilibrium model). Using the standard definition of the simplex $\Delta_d := \left\{ z \in \mathbb{R}^d \mid z \geq 0, \sum_{i=1}^d z_i = 1 \right\}$, consider an economy where there are n goods with relative prices $y \in \Delta_n$, which can be produced by m processes with relative intensities $x \in \Delta_m$. Let $R_{ij} \geq 0$ and $S_{ij} \geq 0$ denote the number of units of the jth good produced and consumed by the ith process, respectively. By [61, §5], a pair $(x^\star, y^\star)$ for which this economy is expanding corresponds to a minimax solution of

$$\underset{x \in \Delta_m}{\text{minimize}} \; \underset{y \in \Delta_n}{\text{maximize}} \; f(x, y) := \frac{\langle x, Ry \rangle}{\langle x, Sy \rangle}. \tag{2.19}$$

Defining $z := (x, y)$, the problem of finding stationary points of (2.19) corresponds to finding a zero of $T(z) := F(z) + N_C(z)$, where $F(z) = \begin{pmatrix} \nabla_x f(x,y) \\ -\nabla_y f(x,y) \end{pmatrix}$ and $C = \Delta_2 \times \Delta_2$. Let $\epsilon \in (0, 1)$ and let

$$R = \begin{bmatrix} 0 & 1 + \epsilon/2 \\ 2 - \epsilon/2 & 2 \end{bmatrix} \quad \text{and} \quad S = \begin{bmatrix} 1/2 & 1/2 \\ 1 & 1 \end{bmatrix}. \tag{2.20}$$

Then, the unique Nash equilibrium $z^\star := (x^\star, y^\star) = [\, 0 \; 1 \; 0 \; 1 \,]^\top$ of (2.19) is a $\left(\frac{\epsilon-1}{4}\right)$-weak Minty solution of T. A numerical experiment leveraging this result can be found in Figure 4. $\square$

2.2 Last-iterate convergence of PPPA under comonotonicity

The convergence result from Theorem 2.4*(iv)* is expressed in terms of a best-so-far $\mathcal{O}(1/(N+1))$ rate for $(q_P(z^k - \bar{z}^k))_{k \in \mathbb{N}}$. To obtain stronger last-iterate convergence rates, governing inequality (2.1) needs to hold between any

two points in the graph of T instead of only between any point in the graph of T and (a subset of) zer T. This observation leads us to introduce the class of V-comonotone operators.

Definition 2.8 (V-comonotone operator). *An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be V-comonotone for some symmetric matrix $V \in \mathbb{S}^n$ if*

$$\langle x - x', y - y' \rangle \geq q_V(y - y'), \qquad \text{for all } (x, y), (x', y') \in \text{gph } T. \tag{2.21}$$

It is said to be maximally V-comonotone if its graph is not strictly contained in the graph of another V-comonotone operator.

This definition naturally reduces to the notion of (maximally) ρ-comonotone operators when $V = \rho I$. The class of operators with V-oblique weak Minty solutions is significantly larger than the class of (maximally) V-comonotone operators. For instance, the following proposition establishes that the class of maximally V-comonotone operator only consist of operators which do not have isolated zeros.

Proposition 2.9. *If an operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is maximally V-comonotone, then* zer T *is convex.*

Proof. First, observe that zer $(T^{-1} + V)^{-1} = (T^{-1} + V)(0) = $ zer T. Furthermore, note that by definition $T^{-1} + V$ is maximally monotone, and thus so is $(T^{-1} + V)^{-1}$. Since the set of zeros of a maximally monotone operator is convex [9, Prop. 23.39], the claim is established. $\qquad \square$

For maximally V-comonotone operators, the convergence of PPPA depends merely upon the selection of the preconditioner P and the relaxation parameters λ_k. This is a consequence of the following proposition, which allows to directly verify the underlying assumptions of Theorem 2.4.

Proposition 2.10. *Suppose that T is maximally V-comonotone, that* zer T *is nonempty and that P is selected according to Assumption I.A4. Then, Assumption I holds.*

Proof. Let $(x^k, y^k) \in \text{gph } T$, where $x^k \to \bar{x}$ and $y^k \to \bar{y}$. Note that $x^k \in (T^{-1} - V)^{-1}(y^k - Vx^k)$. Since T is maximally V-comonotone, $(T^{-1} - V)^{-1}$ is maximally monotone and thus also outer semicontinuous [77, Ex. 12.8]. Consequently, it follows that $\bar{x} \in (T^{-1} - V)^{-1}(\bar{y} - V\bar{x})$. This implies that $\bar{x} \in T\bar{y}$, showing outer semicontinuity of T. Assumption I.A2 follows directly from Proposition A.2 and Assumption I.A3 holds by definition of V-comonotonicity and using that zer T is nonempty. $\qquad \square$

Leveraging Proposition 2.10, sublinear convergence rates can be obtained for the sequence $(q_P(z^k - \bar{z}^k))_{k \in \mathbb{N}}$ in the maximally V-comonotone setting by showing that this sequence is monotonically nonincreasing and using the best-iterate result from Theorem 2.4(iv), as detailed below.

Theorem 2.11 (last-iterate convergence). *Suppose that T is maximally V-comonotone, that* zer T *is nonempty and that P is selected according to Assumption I.A4. Consider a sequence $(z^k, \bar{z}^k)_{k \in \mathbb{N}}$ generated by PPPA starting from $z^0 \in \mathbb{R}^n$ with relaxation parameters $\lambda_k \in (0, 2\eta_{\min})$ such that $\lambda_k(2\eta_{\min} - \lambda_k) \geq \kappa > 0$ uniformly for all k. Then, for all $z^\star \in$ zer T, the following convergence estimates hold:*

$$q_P(z^N - \bar{z}^N) \leq \frac{q_P(z^0 - z^\star)}{(N + 1)\kappa} \quad and \quad q_P(z^N - \bar{z}^N) = o(1/(N+1)). \tag{2.22}$$

Proof. By update rule (PPPA) it holds that

$$P(z^k - \bar{z}^k) \in T\bar{z}^k \quad \text{and} \quad P(z^{k+1} - \bar{z}^{k+1}) \in T\bar{z}^{k+1}.$$

Defining $\tilde{z}^k := z^k - \bar{z}^k$ and using V-comonotonicity of T between the points $\bar{z}^k$ and $\bar{z}^{k+1}$, this implies that

$$\langle P(\tilde{z}^k - \tilde{z}^{k+1}), z^k - z^{k+1} \rangle \geq q_{P+PVP}(\tilde{z}^k - \tilde{z}^{k+1}),$$

which by update rule (2.13) reduces to

$$\langle 2P(\tilde{z}^k - \tilde{z}^{k+1}), \tilde{z}^k \rangle \geq \tfrac{2}{\lambda_k} q_{P+PVP}(\tilde{z}^k - \tilde{z}^{k+1}).$$

12

Since $q_P(u) - q_P(v) = \langle 2Pu, u - v \rangle - q_P(u - v)$ for any $u, v \in \mathbb{R}^n$ and using that $P = UXU^\top$, this implies that

$$\begin{aligned}
q_P(\bar{z}^k) - q_P(\bar{z}^{k+1}) &= \langle 2P\bar{z}^k, \bar{z}^k - \bar{z}^{k+1} \rangle - q_P(\bar{z}^k - \bar{z}^{k+1}) \\
&\geq \tfrac{2}{\lambda_k} q_{P+PVP}(\bar{z}^k - \bar{z}^{k+1}) - q_P(\bar{z}^k - \bar{z}^{k+1}) \\
&= \tfrac{2}{\lambda_k} q_{UX^{1/2}(I+X^{1/2}U^\top VUX^{1/2})X^{1/2}U^\top}(\bar{z}^k - \bar{z}^{k+1}) - q_P(\bar{z}^k - \bar{z}^{k+1}) \\
(2.3) &\geq \left(\tfrac{2\eta_{\min}}{\lambda_k} - 1 \right) q_P(\bar{z}^k - \bar{z}^{k+1}),
\end{aligned}$$

establishing that $q_P(z^k - \bar{z}^k)$ is nonincreasing. Combining this with Theorem 2.4*(iv)* yields the claimed big-$\mathcal{O}$ convergence rate. Finally, the little-o convergence follows from [23, Lem. 3-(1a)]. $\qquad\square$

Theorem 2.11 provides a general framework for the sublinear convergence rate of PPPA, unifying many existing results in literature. Among others, we retrieve the big-$\mathcal{O}$ rates for standard PPA (where $P = \gamma^{-1}I$ and $\lambda_k = 1$) in the monotone setting (where $V = 0$) [14, Prop. 8] and the weak Minty setting (where $V = \rho I$) [39, Thm. 3.1], as well as the rates obtained for PPPA in the monotone setting [49, Thm. 3.4]. Moreover, it is worth highlighting that when $P = \gamma^{-1}I$, Theorem 2.11 indicates that the optimal rate is obtained for $\lambda_k = \eta_{\min}$, as opposed to the usual choice of $\lambda_k = 1$.

2.3 Linear convergence of PPPA under metric subregularity

Metric subregularity is a well-known concept used to describe the behavior of a set-valued mapping in a neighborhood around a given point and can be seen as a one-point version of metric regularity. Many algorithms used in optimization, such as splitting methods, have been shown to have good convergence properties when applied to operators that are metrically subregular, see e.g. [86, 27, 49, 48, 36]. Following [25, Ex. 3H.4], we define metric subregularity as follows.

Definition 2.12 (metric subregularity)**.** *A set-valued mapping* $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^d$ *is metrically subregular at* $\bar{x}$ *for* $\bar{y}$ *if* $(\bar{x}, \bar{y}) \in \operatorname{gph} T$ *and there exists a positive constant* ζ *together with a neighborhood of subregularity* $\mathscr{U}$ *of* $\bar{x}$ *such that*

$$\operatorname{dist}(x, T^{-1}\bar{y}) \leq \zeta \operatorname{dist}(\bar{y}, Tx), \qquad \text{for all } x \in \mathscr{U}. \tag{2.23}$$

There exist many prominent examples of operator classes that are metrically subregular. For instance, all piecewise polyhedral mappings are metrically subregular at all points in their graph [25, §3]. Another important example is the subdifferential operator ∂f of a proper lsc convex function satisfying quadratic growth conditions (cf. [4]). We refer the interested reader to [25, §3] and [77, §9] for further discussion.

The main result of this subsection is Theorem 2.13, where it is shown that local linear convergence of PPPA can be attained under an additional metric subregularity assumption. If the metric subregularity holds globally, i.e., $\mathscr{U} = \mathbb{R}^n$ in Definition 2.12, then the linear convergence holds globally. This result will in turn lead to sufficient conditions for (local) linear convergence of DRS for (possibly) nonmonotone piecewise polyhedral mappings (cf. Theorem 3.6).

Interestingly, when $P = \gamma^{-1}I$ and $V = 0$, Theorem 2.13 establishes that the sequence $(\operatorname{dist}(z^k, \operatorname{zer} T))_{k \in \mathbb{N}}$ converges Q-linearly to zero with the same factor as the one obtained in the maximally monotone setting [79, Thm. 3.2(b)]. For a more detailed discussion on linear convergence of PPA in the monotone setting, we refer the interested reader to [5, 51] and the references therein.

Theorem 2.13 (linear convergence of PPPA under metric subregularity)**.** *In addition to Assumption I, suppose that* $\operatorname{zer} T$ *is nonempty, closed and convex, that Assumption I.*A_3 *holds with* $\mathcal{S}^\star = \operatorname{zer} T$, *that* $(P + T)^{-1} \circ P$ *is continuous, and that* T *is metrically subregular with modulus* $\zeta \in \mathbb{R}_+$ *at all* $z^\star \in \operatorname{zer} T$ *for* 0. *Consider a sequence* $(z^k, \bar{z}^k)_{k \in \mathbb{N}}$ *generated by PPPA starting from* $z^0 \in \mathbb{R}^n$ *with relaxation parameters* $\lambda_k \in (0, 2\alpha_k)$ *such that* $\lambda_k(2\alpha_k - \lambda_k) \geq \kappa > 0$ *for all* k *large enough, where* α_k *is defined as in* (2.12). *Then,* $(\operatorname{dist}_Q(\Pi_{\mathcal{R}(P)} z^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T))_{k \in \mathbb{N}}$ *converges* Q-linearly to zero with Q-factor $\sqrt{1 - \kappa/(1+\zeta\|P\|)^2}$, $(\Pi_{\mathcal{R}(P)} z^k)_{k \in \mathbb{N}}$ *converges* R-linearly to some element of $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$ *and* $q_P(z^k - \bar{z}^k)$ *converges* R-linearly to zero.

Proof. By Theorem 2.4*(v)* the sequence $(\bar{z}^k)_{k \in \mathbb{N}}$ converges to a point $z^\star \in \operatorname{zer} T$. Up to possibly discarding initial iterates, it holds that $\bar{z}^k \in \mathscr{U}_{z^\star}$, where $\mathscr{U}_{z^\star}$ denotes the neighborhood of subregularity associated with $z^\star$. Metric subregularity of T at $z^\star$ for 0 along with the fact that $P(z^k - \bar{z}^k) \in T\bar{z}^k$ (cf. Theorem 2.4*(i)*) implies that

$$\operatorname{dist}(\bar{z}^k, \operatorname{zer} T) \leq \zeta \|P(z^k - \bar{z}^k)\|, \tag{2.24}$$

13

for some $\zeta \in (0, \infty)$. The distance of the shadow sequence from the set $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$ may therefore be bounded as

$$\operatorname{dist}(\bar{w}^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) = \inf_{v \in \Pi_{\mathcal{R}(P)} \operatorname{zer} T} \{\|v - \bar{w}^k\|\} = \inf_{v \in \operatorname{zer} T} \{\|\Pi_{\mathcal{R}(P)}(v - \bar{z}^k)\|\} \le \|\Pi_{\mathcal{R}(P)}\| \operatorname{dist}(\bar{z}^k, \operatorname{zer} T)$$

$$\overset{(2.24)}{\le} \zeta \|P(z^k - \bar{z}^k)\| \tag{2.25}$$

where we used that $\|\Pi_{\mathcal{R}(P)}\| = 1$. Note that since the set $\mathcal{S}^\star = \operatorname{zer} T$ is convex, so is $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$, implying that $\operatorname{dist}(\bar{w}^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) = \|\bar{w}^k - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}(\bar{w}^k)\|$ is attained. By the triangle inequality

$$\|w^k - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}(\bar{w}^k)\| \le \|\bar{w}^k - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}(\bar{w}^k)\| + \|w^k - \bar{w}^k\| = \operatorname{dist}(\bar{w}^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) + \|w^k - \bar{w}^k\|$$

$$\overset{(2.25)}{\le} \zeta \|P(z^k - \bar{z}^k)\| + \|w^k - \bar{w}^k\|$$

$$\overset{(2.16)}{\le} \xi \|w^k - \bar{w}^k\|_Q,$$

where $\xi = \zeta \|P\|^{1/2} + \|P\|^{-1/2}$. It then follows that

$$\operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) = \inf_{v \in \Pi_{\mathcal{R}(P)} \operatorname{zer} T} \{\|w^k - v\|_Q^2\} \le \|w^k - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}(\bar{w}^k)\|_Q^2 \le \xi^2 \|P\| \|w^k - \bar{w}^k\|_Q^2. \tag{2.26}$$

Consequently, due to (2.14) and (2.26) we obtain for all k large enough that

$$\operatorname{dist}_Q^2(w^{k+1}, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) \le \|w^{k+1} - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}^Q(w^k)\|_Q^2$$

$$\overset{(2.14)}{\le} \|w^k - \Pi_{\Pi_{\mathcal{R}(P)} \operatorname{zer} T}^Q(w^k)\|_Q^2 - \lambda_k(2\alpha_k - \lambda_k)\|w^k - \bar{w}^k\|_Q^2$$

$$= \operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) - \lambda_k(2\alpha_k - \lambda_k)\|w^k - \bar{w}^k\|_Q^2 \tag{2.27}$$

$$\overset{(2.26)}{\le} \left(1 - \frac{\kappa}{\xi^2 \|P\|}\right) \operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T),$$

establishing that the sequence $(\operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T))_{k \in \mathbb{N}}$ converges Q-linearly to zero. Therefore, it follows from (2.27) that $(\|w^k - \bar{w}^k\|_Q)_{k \in \mathbb{N}}$ converges R-linearly to zero since

$$\operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T) \ge \lambda_k(2\alpha_k - \lambda_k)\|w^k - \bar{w}^k\|_Q^2 \ge \kappa \|w^k - \bar{w}^k\|_Q^2.$$

Consequently, by (2.13) it follows that

$$\|w^{k+1} - w^k\|_Q^2 = \lambda_k^2 \|\bar{w}^k - w^k\|_Q^2 \le \frac{\lambda_k^2}{\kappa} \operatorname{dist}_Q^2(w^k, \Pi_{\mathcal{R}(P)} \operatorname{zer} T).$$

Combined with the fact that $\lambda_k < 2\alpha_k < +\infty$ is bounded (the latter inequality follows directly from (2.12)), and since the distance on the right-hand side converges Q-linearly to zero, $(\|w^{k+1} - w^k\|_Q)_{k \in \mathbb{N}}$ converges R-linearly to zero. This implies that $(\|w^k - w^\star\|_Q)_{k \in \mathbb{N}}$ (with $w^\star \in \Pi_{\mathcal{R}(P)} \operatorname{zer} T$ being the point where the sequence converges owing to Theorem 2.4(v)) also converges R-linearly to zero. Noting that $\xi^2 \|P\| = (1 + \zeta \|P\|)^2$ and that Q is positive definite completes the proof. $\qquad \square$

3 Douglas-Rachford splitting

In the previous section, convergence results of the preconditioned proximal point method were provided for a class of nonmonotone operators. In this section, convergence results for DRS are obtained in the nonmonotone setting by leveraging the equivalence between DRS and PPPA. In particular, we provide convergence results whenever the primal-dual operator has oblique weak Minty solution(s), as well as in the more restrictive, comonotone setting, and we show (local) linear convergence for piecewise polyhedral mappings.

3.1 Equivalence between DRS and PPPA

In the monotone setting, the convergence of DRS has been studied in [20] by establishing its connection with the preconditioned proximal point method with a positive semidefinite preconditioner. In the nonmonotone setting, this equivalence heavily relies upon the abstract duality framework introduced in [6], [7, §6.9]. In

this framework, the inclusion problem (P-I) is referred to as the *primal inclusion*. Associated with the primal inclusion is the *dual inclusion*, given by

$$\text{find} \quad y \in \mathbb{R}^n \quad \text{such that} \quad 0 \in T_{\mathrm{D}} y := -A^{-1}(-y) + B^{-1}(y). \tag{D-I}$$

A key property of the dual inclusion is that the primal inclusion is solvable if and only if the dual inclusion is solvable [7, Prop. 6.9.1]. Finally, the *primal-dual inclusion* is given by

$$\text{find} \quad z = (x, y) \in \mathbb{R}^{2n} \quad \text{such that} \quad \begin{bmatrix} 0 \\ 0 \end{bmatrix} \in T_{\mathrm{PD}} z := \begin{bmatrix} Ax \\ B^{-1}y \end{bmatrix} + \begin{bmatrix} y \\ -x \end{bmatrix}. \tag{PD-I}$$

For the primal-dual inclusion, it holds by [7, Prop. 6.9.2] that

$$(x^\star, y^\star) \in \mathrm{zer}\, T_{\mathrm{PD}} \iff (x^\star, -y^\star) \in \mathrm{gph}\, A, \; (x^\star, y^\star) \in \mathrm{gph}\, B,$$

where $\mathrm{zer}\, T_{\mathrm{P}} = \{x^\star \mid \exists y^\star : (x^\star, y^\star) \in \mathrm{zer}\, T_{\mathrm{PD}}\}$ and $\mathrm{zer}\, T_{\mathrm{D}} = \{y^\star \mid \exists x^\star : (x^\star, y^\star) \in \mathrm{zer}\, T_{\mathrm{PD}}\}$. Therefore, the primal-dual inclusion encompasses the solutions of both the primal and the dual inclusion, in the sense that $(x^\star, y^\star)$ is a solution of (PD-I) if and only if $x^\star$ is a solution of (P-I) and $y^\star$ is a solution of (D-I). Hence, to obtain a zero of T_{P}, one may apply PPPA to the primal-dual operator T_{PD} starting from an initial point $z^0 = (x^0, y^0)$ with iterates $z^k = (x^k, y^k)$ and $\bar{z}^k = (\bar{x}^k, \bar{y}^k)$. Most notably, PPPA applied to the primal-dual operator T_{PD} reduces to DRS when the preconditioner P is selected to be equal to

$$P = \begin{bmatrix} \gamma^{-1} \mathrm{I}_n & -\mathrm{I}_n \\ -\mathrm{I}_n & \gamma \mathrm{I}_n \end{bmatrix}, \tag{3.1}$$

where $\gamma \in \mathbb{R}_{++}$ is a strictly positive stepsize parameter. In the following lemma, the equivalence between PPPA and DRS is formally summarized.

Lemma 3.1 (equivalence of DRS and PPPA). *Let $z^0 = (x^0, y^0) \in \mathbb{R}^{2n}$ and set $s^0 = x^0 - \gamma y^0$. Then, to any sequence $(u^k, v^k, s^k)_{k \in \mathbb{N}}$ generated by DRS (initialized with s^0) for solving (P-I), there correspond sequences $(z^k)_{k \in \mathbb{N}} = (x^k, y^k)_{k \in \mathbb{N}}$, $(\bar{z}^k)_{k \in \mathbb{N}} = (\bar{x}^k, \bar{y}^k)_{k \in \mathbb{N}}$ generated by PPPA (initialized with z^0) for solving (PD-I) with preconditioner (3.1) (and vice versa), and the correspondence is as follows.*

(i) $s^k = x^k - \gamma y^k$.

(ii) $\bar{x}^k = u^k$.

(iii) $\bar{y}^k = 1/\gamma(2u^k - s^k - v^k)$.

Proof. See Appendix B.1. ∎

3.2 Convergence analysis of DRS under oblique weak Minty

Despite the fact that DRS is a primal algorithm, Lemma 3.1 shows that it can be interpreted as applying PPPA to the associated primal-dual operator T_{PD}. Based on this equivalence, we proceed to derive convergence results for DRS, leveraging the results obtained in Section 2. First, the underlying assumptions for PPPA provided in Assumption I are translated to conditions on the individual operators A and B as follows.

Assumption II. *In problem (P-I), the following hold.*

A1 *Operators A and B are outer semicontinuous.*

A2 *For the selected stepsizes, the resolvents have full domain, i.e., $\mathrm{dom}\, J_{\gamma A} = \mathrm{dom}\, J_{\gamma B} = \mathbb{R}^n$.*

A3 *There exists a nonempty set $\mathcal{S}^\star \subseteq \mathrm{zer}\, T_{\mathrm{PD}}$ such that the primal-dual operator (PD-I) has V-oblique weak Minty solutions at $\mathcal{S}^\star$, where*

$$V := \mathrm{blkdiag}(\beta_{\mathrm{P}} \mathrm{I}_n, \beta_{\mathrm{D}} \mathrm{I}_n) \tag{3.2}$$

and $\beta_{\mathrm{P}}, \beta_{\mathrm{D}}, \in \mathbb{R}$ satisfy $[\beta_{\mathrm{D}}]_- [\beta_{\mathrm{P}}]_- < 1/4$.

15

Table 2: *Table summarizing which operators from the abstract duality framework satisfy the Minty variational inequality (MVI), depending on the sign of β_P and β_D.*

	$\beta_D \geq 0$	$\beta_D < 0$
$\beta_P \geq 0$	T_P, T_D, T_{PD}	T_P
$\beta_P < 0$	T_D	-

Note that these assumptions are not very restrictive. For instance, only the existence of a single oblique weak Minty solution is required for the primal-dual inclusion. Furthermore, it can be seen that the range of β_P and β_D in Assumption II.A3 covers cases where neither the primal inclusion (P-I) nor the dual inclusion (D-I) satisfies the Minty variational inequality (MVI), which is an immediate consequence of the following lemma (see also Table 2).

Lemma 3.2 (weak Minty for primal and dual operator). *Suppose that Assumption II.A3 holds and let*

$$\mathcal{S}_P^\star := \{x^\star \mid \exists y^\star : (x^\star, y^\star) \in \mathcal{S}^\star\} \quad and \quad \mathcal{S}_D^\star := \{y^\star \mid \exists x^\star : (x^\star, y^\star) \in \mathcal{S}^\star\}.$$

Then, the primal operator T_P has β_P-weak Minty solutions at $\mathcal{S}_P^\star$ and the dual operator T_D has β_D-weak Minty solutions at $\mathcal{S}_D^\star$.

Proof. See Appendix B.1. □

The converse of Lemma 3.2 does not hold. For instance, consider the linear operators $A = -I_n$ and $B = I_n$: in this case both the primal operator T_P and the dual operator T_D reduce to the zero mapping, which has 0-weak Minty solutions everywhere, while the primal-dual operator reduces to $T_{PD} = \begin{bmatrix} -I_n & I_n \\ -I_n & I_n \end{bmatrix}$, which does not have any 0-weak Minty solutions.

The following theorem follows as a consequence of Theorem 2.4 and characterizes the convergence of DRS in the nonmonotone setting, assuming the corresponding primal-dual operator has V-oblique weak Minty solutions. To prove this result, we rely upon Lemma A.5, which shows that when Assumption II holds for DRS, Assumptions I.A1 to I.A3 are satisfied for the corresponding preconditioned proximal point algorithm. Note that Assumption I.A4 is not covered by this lemma. To ensure satisfaction of this final assumption, the stepsize γ of DRS needs to be selected appropriately (see condition (3.3)).

Theorem 3.3 (convergence of DRS). *Suppose that Assumption II holds and that γ lies within the interval*

$$\left(\frac{2[-\beta_P]_+}{1 + \sqrt{1 - 4\beta_D\beta_P}}, \frac{1 + \sqrt{1 - 4\beta_D\beta_P}}{2[-\beta_D]_+} \right) \tag{3.3}$$

(which is nonempty since $[\beta_D]_-[\beta_P]_- < 1/4$). Consider a sequence $(u^k, v^k, s^k)_{k\in\mathbb{N}}$ generated by DRS starting from $s^0 \in \mathbb{R}^n$ with stepsize γ and relaxation parameters $\lambda_k \in (0, 2\alpha)$ such that $\liminf_{k\to\infty} \lambda_k(2\alpha - \lambda_k) > 0$, where $\alpha := 1 + \frac{1}{\gamma}\beta_P + \gamma\beta_D > 0$. Then, either a point for which $u^k = v^k$ is reached in a finite number of iterations, implying that $(u^k, 1/\gamma(u^k - s^k)) \in \text{zer } T_{PD}$, or the following hold for the sequence $(u^k, v^k, s^k)_{k\in\mathbb{N}}$.

(i) *$\gamma^{-1}(u^k - v^k) \in Au^k + Bv^k$ for all k, and $(u^k - v^k)_{k\in\mathbb{N}}$ converges to zero.*

(ii) *Every limit point (if any) of $(u^k, 1/\gamma(u^k - s^k))_{k\in\mathbb{N}}$ belongs to $\text{zer } T_{PD}$.*

(iii) *The sequence $(s^k)_{k\in\mathbb{N}}$ is bounded and its limit points belong to $\{(x^\star - \gamma y^\star) \mid (x^\star, y^\star) \in \text{zer } T_{PD}\}$.*

(iv) *If $\lambda_k(2\alpha - \lambda_k) \geq \kappa > 0$ uniformly for all k, then, for all $s^\star \in \{(x^\star - \gamma y^\star) \mid (x^\star, y^\star) \in \mathcal{S}^\star\}$ we have*

$$\min_{k=0,1,\dots,N} \|u^k - v^k\|^2 \leq \frac{\|s^0 - s^\star\|^2}{(N+1)\kappa}.$$

(v) *If in Assumption II.A3 $\Pi_{\mathcal{R}(P)} \mathcal{S}^\star = \Pi_{\mathcal{R}(P)} \text{zer } T_{PD}$, then $(s^k)_{k\in\mathbb{N}}$ converges to some element of $\{(x^\star - \gamma y^\star) \mid (x^\star, y^\star) \in \text{zer } T_{PD}\}$. If additionally $J_{\gamma A}$ and $J_{\gamma B}$ are (single-valued) continuous, then $(u^k, 1/\gamma(u^k - s^k))_{k\in\mathbb{N}}$ converges to some element of $\text{zer } T_{PD}$.*

16

Proof. Recall that DRS is cast as an instance of PPPA, as summarized in Lemma 3.1. Therefore, all the assertions follow from those of Theorem 2.4 applied to the primal-dual inclusion (PD-I) with P as defined in (3.1). Moreover, by Lemma A.5 it follows immediately that Assumptions I.A1 to I.A3 hold so that it only remains to show that Assumption I.A4 holds whenever γ satisfies (3.3).

Throughout the proof, T_{PD} is as in (PD-I). The preconditioning matrix P has two distinct eigenvalues, namely zero and $\gamma + \gamma^{-1}$, each with a multiplicity of n. Therefore, $P = (\gamma + \gamma^{-1})UU^{\top}$, where U is an orthonormal basis for the range of P as defined in Assumption I.A4. In particular, we have

$$U = \tfrac{1}{\sqrt{\gamma^2+1}}\begin{bmatrix} \mathrm{I} \\ -\gamma\mathrm{I} \end{bmatrix} \quad \text{and} \quad \Pi_{\mathcal{R}(P)} = UU^{\top} = \tfrac{1}{\gamma^2+1}\begin{bmatrix} \mathrm{I} & -\gamma\mathrm{I} \\ -\gamma\mathrm{I} & \gamma^2\mathrm{I} \end{bmatrix}.$$

Therefore, substituting $V := \operatorname{blkdiag}(\beta_{\mathrm{P}}\mathrm{I}_n, \beta_{\mathrm{D}}\mathrm{I}_n)$ into (2.2), it follows that $\eta_{\min} = 1 + \tfrac{1}{\gamma}\beta_{\mathrm{P}} + \gamma\beta_{\mathrm{D}}$. Owing to Fact A.3*(ii)*, $\eta_{\min} > 0$ iff γ satisfies (3.3), and therefore Assumption I.A4 holds.

Having established that the underlying assumptions of Theorem 2.4 are satisfied, it remains to translate the sequences therein in terms of the variables of DRS. Leveraging the equivalences established in Lemma 3.1, it follows that

$$P(z^k - \bar{z}^k) = \left(-\tfrac{1}{\gamma}(u^k - v^k), u^k - v^k \right) \tag{3.4}$$

$$\mathsf{q}_P(z^k - \bar{z}^k) = \tfrac{1}{\gamma}\|u^k - v^k\|^2 \tag{3.5}$$

$$\mathsf{q}_V(P(z^k - \bar{z}^k)) = (\tfrac{1}{\gamma^2}\beta_{\mathrm{P}} + \beta_{\mathrm{D}})\|u^k - v^k\|^2. \tag{3.6}$$

Analogous to the proof of Theorem 2.4, consider the shadow sequences $(w^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} x^k)_{k\in\mathbb{N}}$ and $(\bar{w}^k)_{k\in\mathbb{N}} := (\Pi_{\mathcal{R}(P)} \bar{x}^k)_{k\in\mathbb{N}}$ with update rule (2.13). Owing to Lemma 2.3*(ii)* and (3.4), it holds that $w^k \in \mathscr{D}_{w^k,\bar{w}^k}$ if and only if $u^k = v^k$ Consequently, if $w^k \in \mathscr{D}_{w^k,\bar{w}^k}$ for some $k \in \mathbb{N}$ then the algorithm has reached a point for which $u^k = v^k$ in a finite number of iterations, which by update rule (DRS) immediately implies that $(u^k, {}^1/_\gamma(u^k - s^k)) \in \operatorname{zer} T_{\mathrm{PD}}$. We will consider the case when $w^k \notin \mathscr{D}_{w^k,\bar{w}^k}$ in the remainder of the proof, in which case α_k as defined in (2.12) coincides with $\alpha > 0$ as defined in this theorem.

♠ 3.3*(i)*: By Theorem 2.4*(i)* and (3.4), the sequence $(u^k - v^k)_{k\in\mathbb{N}}$ converges to zero and

$$\tfrac{1}{\gamma}(u^k - v^k) \in A\bar{x}^k + \bar{y}^k, \quad -(u^k - v^k) \in B^{-1}\bar{y}^k - \bar{x}^k,$$

implying the claimed inclusion.

♠ 3.3*(ii)*: By Lemma 3.1, it holds that $(\bar{x}^k, \bar{y}^k)_{k\in\mathbb{N}} = (u^k, {}^1/_\gamma(2u^k - s^k - v^k))_{k\in\mathbb{N}}$. Consequently, it follows from Theorem 2.4*(ii)* that every limit point (if any) of $(u^k, {}^1/_\gamma(2u^k - s^k - v^k))_{k\in\mathbb{N}}$ belongs to $\operatorname{zer} T_{\mathrm{PD}}$. The claimed result then follows from the fact that $(u^k - v^k)_{k\in\mathbb{N}}$ converges to zero.

♠ 3.3*(iii)*: The claim follows directly from Theorem 2.4*(iii)* after noting that $\Pi_{\mathcal{R}(P)} z^k = \tfrac{1}{1+\gamma^2}(s^k, -\gamma s^k)$.

♠ 3.3*(iv)*: The claimed rate follows from Theorem 2.4*(iv)* since $\|P(z^k - \bar{z}^k)\|^2 = (1 + \tfrac{1}{\gamma^2})\|u^k - v^k\|^2$, $\mathsf{q}_P(z^0 - \bar{z}^\star) = \tfrac{1}{\gamma}\|s^0 - s^\star\|^2$ and $\|P\| = \gamma + \tfrac{1}{\gamma}$.

♠ 3.3*(v)*: The assertion for $(s^k)_{k\in\mathbb{N}}$ is of immediate verification based on the assertion for $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ in Theorem 2.4*(v)*. Furthermore, by Lemma 3.1, we know that $\bar{z}^k \in (P + T_{\mathrm{PD}})^{-1}Pz^k$ iff $\bar{x}^k \in J_{\gamma A}(x^k - \gamma y^k)$ and $\bar{y}^k \in {}^1/_\gamma(2J_{\gamma A}(x^k - \gamma y^k) - x^k + \gamma y^k - J_{\gamma B}(2J_{\gamma A}(x^k - \gamma y^k) - x^k + \gamma y^k))$. Consequently, if $J_{\gamma A}$ and $J_{\gamma B}$ are (single-valued) continuous, then $(P + T_{\mathrm{PD}})^{-1}P$ is (single-valued) continuous as well. Therefore, the claim for $(u^k, {}^1/_\gamma(u^k - s^k))_{k\in\mathbb{N}}$ follows from the assertion for $(\bar{x}^k, \bar{y}^k)_{k\in\mathbb{N}} = (u^k, {}^1/_\gamma(2u^k - s^k - v^k))_{k\in\mathbb{N}}$ in Theorem 2.4*(v)* and the fact that $(u^k - v^k)_{k\in\mathbb{N}}$ converges to zero. □

Note that in Theorem 3.3*(v)* we show that under suitable conditions $(u^k, {}^1/_\gamma(u^k - s^k))_{k\in\mathbb{N}}$ converges to some element of $\operatorname{zer} T_{\mathrm{PD}}$. By [7, Prop. 6.9.2], this immediately implies that $(u^k)_{k\in\mathbb{N}}$ converges to some element of $\operatorname{zer} T_{\mathrm{P}}$ and $({}^1/_\gamma(u^k - s^k))_{k\in\mathbb{N}}$ converges to some element of $\operatorname{zer} T_{\mathrm{D}}$. An analogous argument holds for the limit point result of Theorem 3.3*(ii)*.

Under an additional V-comonotonicity assumption on the primal-dual operator, last-iterate convergence rates can be obtained for the sequence $(\|u^k - v^k\|^2)_{k\in\mathbb{N}}$. This result follows directly from Theorem 2.11 by casting DRS as an instance of PPPA, analogous to the proof of Theorem 3.3.

Corollary 3.4 (last-iterate convergence)**.** *Suppose that the primal-dual operator T_{PD} is maximally $V :=$* blkdiag$(\beta_{\mathrm{P}}\mathrm{I}_n, \beta_{\mathrm{D}}\mathrm{I}_n)$*-comonotone, that $[\beta_{\mathrm{D}}]_-[\beta_{\mathrm{P}}]_- < \frac{1}{4}$, that $\mathrm{zer}\, T_{\mathrm{PD}}$ is nonempty and that γ is selected according to* (3.3)*. Consider a sequence $(u^k, v^k, s^k)_{k\in\mathbb{N}}$ generated by DRS starting from $s^0 \in \mathbb{R}^n$ with stepsize γ and relaxation parameters $\lambda_k \in (0, 2\alpha)$ such that $\lambda_k(2\alpha - \lambda_k) \geq \kappa > 0$ uniformly for all k, where $\alpha := 1 + \frac{1}{\gamma}\beta_{\mathrm{P}} + \gamma\beta_{\mathrm{D}} > 0$. Then, for all $s^\star \in \{(x^\star - \gamma y^\star) \mid (x^\star, y^\star) \in \mathrm{zer}\, T_{\mathrm{PD}}\}$, the following convergence estimates hold:*

$$\|u^N - v^N\|^2 \leq \frac{\|s^0 - s^\star\|^2}{(N+1)\kappa} \quad and \quad \|u^N - v^N\|^2 = o(1/(N+1)).$$

Finally, consider the class of piecewise polyhedral mappings, which trivially includes piecewise linear mappings and the subgradient of all piecewise linear quadratic convex functions such as normal cones of polyhedral sets [77, §12].

Definition 3.5 (piecewise polyhedral mappings)**.** *A set-valued mapping $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be piecewise polyhedral if its graph can be expressed as the union of finitely many polyhedral sets.*

The following theorem shows that DRS achieves (local) linear convergence for (possibly) nonmonotone piecewise polyhedral mappings. The proof is similar to [48, Lem. IV.4] and relies on the fact that piecewise polyhedral mappings are metrically subregular at all points in their graph [25, §3] and on previous results from Theorem 2.13 and Lemma 3.1.

Theorem 3.6 (linear convergence of DRS for piecewise polyhedral mappings)**.** *Suppose that A and B are piecewise polyhedral mappings, that $\mathrm{zer}\, T_{\mathrm{PD}}$ is nonempty closed convex, and that Assumption II.A3 holds with $\mathcal{S}^\star = \mathrm{zer}\, T_{\mathrm{PD}}$. Set the stepsize γ according to* (3.3) *and suppose that the resolvents $J_{\gamma A}$, $J_{\gamma B}$ have full domain and are continuous. Consider a sequence $(u^k, v^k, s^k)_{k\in\mathbb{N}}$ generated by DRS starting from $s^0 \in \mathbb{R}^n$ with stepsize γ and relaxation parameters $\lambda_k \in (0, 2\alpha)$ such that $\liminf_{k\to\infty} \lambda_k(2\alpha - \lambda_k) > 0$, where α is defined as in Theorem 3.3. Then, $(s^k)_{k\in\mathbb{N}}$ converges R-linearly to some element of $\{(x^\star - \gamma y^\star) \mid (x^\star, y^\star) \in \mathrm{zer}\, T_{\mathrm{PD}}\}$ and $(\|u^k - v^k\|)_{k\in\mathbb{N}}$ converges R-linearly to zero.*

Proof. See Appendix B.1. □

4 Semimonotone operators

In this section, we introduce the class of semimonotone operators and provide several examples of well-known function classes with semimonotone subdifferentials. Sufficient conditions for the convergence of DRS applied to the sum of two semimonotone operators will be provided in Section 5.

We begin by providing the definition of (μ, ρ)-semimonotone operators, which can be seen as a natural extension of (hypo)monotone and co(hypo)monotone operators. The semimonotone terminology adapted here is inspired by the notion of semimonotonicity introduced in [63, Def. 2] (which corresponds to the case $\mu = \rho$).

Definition 4.1 (semimonotonicity)**.** *Let $\mu, \rho \in \mathbb{R}$. An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be (μ, ρ)-semimonotone at $(x', y') \in \mathrm{gph}\, A$ if*

$$\langle x - x', y - y' \rangle \geq \mu\|x - x'\|^2 + \rho\|y - y'\|^2, \qquad for\ all\ (x, y) \in \mathrm{gph}\, A. \tag{4.1}$$

An operator A is said to be (μ, ρ)-semimonotone if it is (μ, ρ)-semimonotone at all $(x', y') \in \mathrm{gph}\, A$. It is said to be maximally (μ, ρ)-semimonotone if its graph is not strictly contained in the graph of another (μ, ρ)-semimonotone operator.

Due to the additional degree of freedom this operator class enjoys, it is able to cover a wide array of other commonly used operator classes available in literature. We list the most prominent connections below.

Remark 4.2 (relationship with other types of operators)**.**

  (i) $(0, 0)$-semimonotonicity is equivalent to the monotonicity of A (and of its inverse A^{-1}).

  (ii) $(\mu, 0)$-semimonotonicity is equivalent to μ-monotonicity, which is also known as $|\mu|$-*hypo*monotonicity [77, Example 12.28] when $\mu < 0$ and as *strong* monotonicity when $\mu > 0$.

18
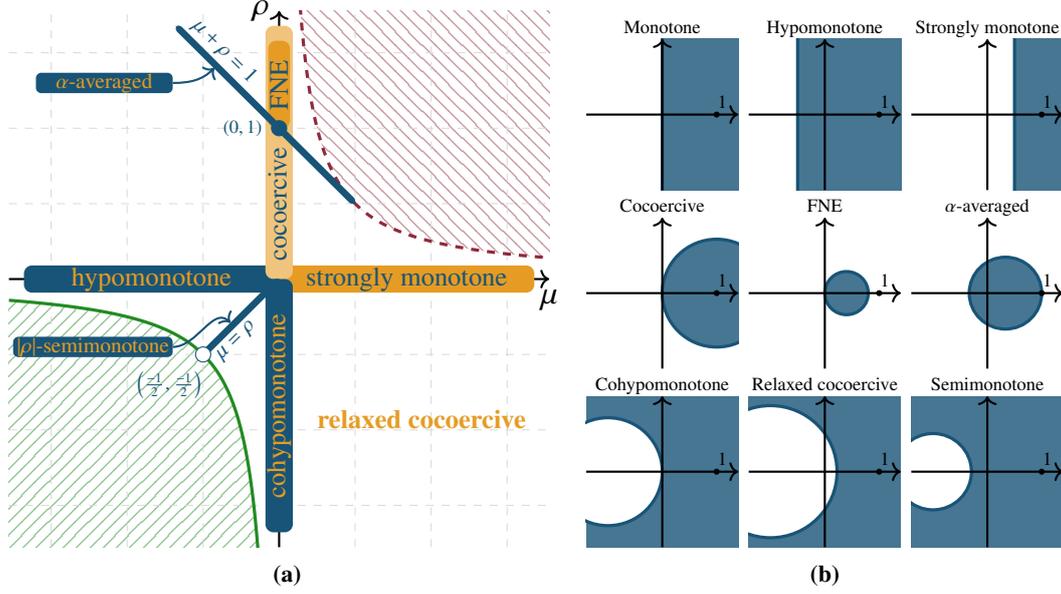
Figure 5: *(left) Relationship between semimonotonicity and other types of operators. The green hatched area represents the region where all operators satisfy (μ, ρ)-semimonotonicity and the red hatched area represents the region for which there do not exist any (μ, ρ)-semimonotone operators (see Proposition 4.3). (right) Scaled Relative Graph (SRG) for different operator classes. For instance, in the top row, the SRG of a μ-monotone operator is visualized [78, Prop. 3.3 and Thm. 3.5]. In the bottom right figure a (μ, ρ)-semimonotone operator with negative μ and ρ is visualized. For more details on the SRG and similar geometric illustrations, we refer the interested reader to [29, 37, 78].*

(iii) Similarly, $(0, \rho)$-semimonotonicity is equivalent to ρ-comonotonicity, which is also referred to as $|\rho|$-co*hypo*monotonicity when $\rho < 0$ and as ρ-cocoercivity when $\rho > 0$.

(iv) $(0, \rho)$-semimonotonicity of an operator T at a point $(x^\star, 0) \in \text{gph}\,T$ is equivalent to T having ρ-weak Minty solutions at $x^\star \in \text{zer}\,T$, sometimes also referred to as ρ-star-cocoercivity around $x^\star$ [38, Def. 2.1].

(v) When $\mu = \rho < 0$, our definition of (μ, ρ)-semimonotonicity is equivalent to the definition of $|\rho|$-semimonotonicity proposed in [63, Def. 2].

(vi) α-averagedness of an operator is equivalent to $(1 - 1/(2-2\alpha), 1/(2-2\alpha))$-semimonotonicity [9, Prop. 4.35(iv)]. Additionally, an operator is firmly nonexpansive if and only if it is $1/2$-averaged [9, Remark 4.34(iii)], which holds for any $(0, \rho)$-semimonotone operator with $\rho \geq 1$.

(vii) When $\mu > 0$ and $\rho < 0$, then the definition of (μ, ρ)-semimonotonicity coincides with the definition of $(-\rho, \mu)$-relaxed-cocoercivity [45, Def. 2.3]. □

Figure 5 provides the Scaled Relative Graph (SRG) [78] of different operator classes and a graphical representation of the connections between them. In Figure 5a, the green (resp. red) hatched area represents the region where all operators (resp. no operators) satisfy (μ, ρ)-semimonotonicity, which is a consequence of the following proposition.

Proposition 4.3 (existence of semimonotone operators)**.** *Let $\mu, \rho \in \mathbb{R}$. The following hold.*

(i) *If $[\mu]_- [\rho]_- \geq \frac{1}{4}$, then all operators $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfy the definition of (μ, ρ)-semimonotonicity.*

(ii) *If $[\mu]_+ [\rho]_+ > \frac{1}{4}$, then there does not exist any operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ that is (μ, ρ)-semimonotone.*

(iii) *If $[\mu]_+ [\rho]_+ = \frac{1}{4}$, then operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$*

  1. is (μ, ρ)-semimonotone iff for some $c \in \mathbb{R}^n$ it holds that $\text{gph}\,A \subseteq \text{gph}\,T$, where $T : x \mapsto 2\mu x + c$.

  2. is maximally (μ, ρ)-semimonotone iff for some $c \in \mathbb{R}^n$ it holds that $A : x \mapsto 2\mu x + c$.

19

Proof. See Appendix B.2. □

We proceed to examine basic properties and calculus rules for the class of (maximally) semimonotone operators. For instance, their inverses belong to the same class of operators, with the roles of μ and ρ reversed. Additionally, the following proposition analyzes the scaling and shifting of semimonotone operators.

Proposition 4.4 (inverting, shifting and scaling). *Let operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (maximally) (μ, ρ)-semimonotone [at $(x', y') \in \text{gph}\, A$]. Then, the following hold.*

(i) *The inverse operator A^{-1} is (maximally) (ρ, μ)-semimonotone [at $(y', x') \in \text{gph}\, A^{-1}$].*

(ii) *For all $\alpha \in \mathbb{R}_{++}$ and $u, w \in \mathbb{R}^n$, operator $T(x) := w + \alpha A(x + u)$ is (maximally) $(\alpha\mu, \alpha^{-1}\rho)$-semimonotone [at $(x' - u, w + \alpha y')$].*

Linear operators are a particular instance of semimonotone operators. For these operators, there exists an interplay between μ and ρ, which is summarized in the following proposition. This result generalizes [10, Prop. 5.1] for μ-monotone and ρ-comonotone operators. Additionally, a subsequent corollary for multiple of identity operators is provided, including its maximality.

Proposition 4.5 (linear operator). *Let $A \in \mathbb{R}^{n \times n}$ and let $\mu, \rho \in \mathbb{R}$. Then, A is (μ, ρ)-semimonotone if and only if $\frac{1}{2}(A + A^\top) - \rho A^\top A \succeq \mu I$.*

Proof. Consider $x \in \mathbb{R}^n$. Then, it holds by (μ, ρ)-semimonotonicity that $\langle x, Ax \rangle \geq \mu\|x\|^2 + \rho\|Ax\|^2 \Leftrightarrow \langle x, (A - \mu I - \rho A^\top A)x \rangle \geq 0 \Leftrightarrow \langle x, \left(\frac{1}{2}(A + A^\top) - \mu I - \rho A^\top A\right)x \rangle \geq 0 \Leftrightarrow \frac{1}{2}(A + A^\top) - \mu I - \rho A^\top A \succeq 0$. □

Corollary 4.6 (multiple of identity). *If $A = \alpha I_n$ for some $\alpha \in \mathbb{R}$, then A is $(\alpha(1 - c\alpha), c)$-semimonotone for any $c \in \mathbb{R}$. Furthermore, A is maximal except when $\alpha < 0$ and $c = 1/2\alpha$.*

Proof. See Appendix B.2. □

The upcoming proposition discusses the sum and the parallel sum of two semimonotone operators, A and B. Notice that the semimonotonicity parameters of the resulting operator always involve both a sum and a parallel sum of A and B's semimonotonicity parameters. As expected, the commutative property holds as swapping A and B does not affect the result.

Proposition 4.7 (sum and parallel sum). *Let operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (μ_A, ρ_A)-semimonotone [at $(x'_A, y'_A) \in \text{gph}\, A$] and operator $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (μ_B, ρ_B)-semimonotone [at $(x'_B, y'_B) \in \text{gph}\, B$]. Then, it holds that*

(i) *[if $x' = x'_A = x'_B$, then] $A + B$ is $(\mu_A + \mu_B, \rho_A \square \rho_B)$-semimonotone [at $(x', y'_A + y'_B)$] if either $\rho_A + \rho_B > 0$ or $\rho_A = \rho_B = 0$.*

(ii) *[if $y' = y'_A = y'_B$, then] $A \square B$ is $(\mu_A \square \mu_B, \rho_A + \rho_B)$-semimonotone [at $(x'_A + x'_B, y')$] if either $\mu_A + \mu_B > 0$ or $\mu_A = \mu_B = 0$.*

Proof. See Appendix B.2. □

Combining the previous proposition for the sum with the characterization of multiple of identity operators from Corollary 4.6, the semimonotonicity of the sum of a semimonotone and a multiple of identity operator is obtained in Proposition 4.8(ii). However, under a particular condition, a stronger equivalence relationship can be obtained by exploiting the multiple of identity structure present. This result is provided in Proposition 4.8(i) and generalizes [10, Lem. 2.6 & 2.8].

Proposition 4.8 (sum with identity). *Let $\mu_A, \rho_A \in \mathbb{R}$. Consider the operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and define $T := A + \alpha I$ where $\alpha \in \mathbb{R}$. Then, the following hold.*

(i) *If $1 + 2\rho_A \alpha > 0$, then T is (maximally) $\left(\frac{\mu_A + \alpha(1 + \rho_A \alpha)}{1 + 2\rho_A \alpha}, \frac{\rho_A}{1 + 2\rho_A \alpha}\right)$-semimonotone [at $(x', y' + \alpha x') \in \text{gph}\, T$] if and only if A is (maximally) (μ_A, ρ_A)-semimonotone [at $(x', y') \in \text{gph}\, A$].*

(ii) *If $1 + 2\rho_A \alpha \leq 0$, then T is $(\mu_A + \alpha(1 - c\alpha), \rho_A \square c)$-semimonotone [at $(x', y' + \alpha x') \in \text{gph}\, T$] for all $c > -\rho_A$ if A is (μ_A, ρ_A)-semimonotone [at $(x', y') \in \text{gph}\, A$].*

Proof. See Appendix B.2. □

Leveraging previous results, we now formalize the relationship between the semimonotonicity of an operator T defined as a function of another (maximally) monotone operator A. In practical applications, this result can be leveraged to check (maximal) semimonotonicity of a given operator or to generate semimonotone operators from a given monotone operator.

Corollary 4.9 (connection to monotone operators). *Let $\xi, \nu \in \mathbb{R}$. Consider the operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and define $T := (A + \xi\mathrm{id})^{-1} + \nu\mathrm{id}$. Then, the following hold.*

(i) If $1 + 2\xi\nu > 0$, then T is (maximally) $\left(\frac{\nu(1+\xi\nu)}{1+2\xi\nu}, \frac{\xi}{1+2\xi\nu}\right)$-semimonotone if and only if A is (maximally) monotone.

(ii) If $1 + 2\xi\nu \leq 0$, then T is $\left(\nu(1 - c\nu), \xi \square c\right)$-semimonotone for all $c > -\xi$ if A is monotone.

Proof. By applying Propositions 4.4(i) and 4.8 it follows that A is (maximally) monotone if and only if $A + \xi\mathrm{id}$ is (maximally) $(\xi, 0)$-semimonotone if and only if $(A+\xi\mathrm{id})^{-1}$ is (maximally) $(0, \xi)$-semimonotone. The claimed result then follows by using Proposition 4.8 again. □

When $\nu = 0$, Corollary 4.9(i) reduces to a known equivalence between comonotonicity of an operator and monotonicity of its Yosida regularization, see [16, Prop. 6.9.3]. The following result is a direct consequence of Corollary 4.9(i), noting that for the choice of ξ and ν from (4.2) it holds that $1 + 2\xi\nu = 1/\sqrt{1-4\rho\mu} > 0$. This result generalizes [63, Thm. 2], which only covers the case $\mu = \rho$.

Corollary 4.10. *A set-valued operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is (maximally) (μ, ρ)-semimonotone with $\rho\mu < \frac{1}{4}$ if and only if $A := (T - \nu\mathrm{id})^{-1} - \xi\mathrm{id}$ is a (maximally) monotone operator where*

$$\xi = \frac{\rho}{\sqrt{1-4\mu\rho}} \quad \text{and} \quad \nu = \frac{2\mu}{1 + \sqrt{1-4\mu\rho}}. \tag{4.2}$$

In light of the above connection with maximally monotone operators, a (μ, ρ)-semimonotone operator A with $\rho\mu < \frac{1}{4}$ is maximal if $(A - \nu\mathrm{id})^{-1}$ is continuous (since any continuous monotone mapping is maximally monotone). Additionally, the following proposition shows that the graph of a maximally (μ, ρ)-semimonotone operator is closed.

Proposition 4.11 (outer semicontinuity of a semimonotone operator). *Let $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a maximally (μ, ρ)-semimonotone operator with either $\rho\mu < \frac{1}{4}$ or $[\mu]_+[\rho]_+ = \frac{1}{4}$. Then, A is outer semicontinuous.*

Proof. See Appendix B.2. □

We conclude this section by showing that if the stepsize is selected appropriately, the resolvent of a semimonotone operator has full domain and is Lipschitz continuous. This result generalizes [10, Prop. 2.17] for comonotone operators and [63, Thm. 3], which is restricted to the case $\mu = \rho$.

Proposition 4.12 (resolvent of a semimonotone operator). *Let $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a (μ, ρ)-semimonotone operator with either $\rho\mu < \frac{1}{4}$ or $[\mu]_+[\rho]_+ = \frac{1}{4}$. Let*

$$\gamma \in \left(\frac{2[-\rho]_+}{1+\sqrt{1-4\rho\mu}}, \frac{1+\sqrt{1-4\rho\mu}}{2[-\mu]_+}\right). \tag{4.3}$$

Then, the following hold.

(i) The resolvent $J_{\gamma A}$ has full domain if and only if A is maximal.

(ii) The resolvent $J_{\gamma A}$ is Lipschitz continuous with constant $L(\gamma) = \frac{|1+2\rho\gamma^{-1}| + \sqrt{1-4\mu\rho}}{2(1+\mu\gamma+\rho\gamma^{-1})}$.

Proof. See Appendix B.2. □

4.1 Examples of functions with semimonotone subdifferentials

In what follows, we present various function classes of which the (limiting) subdifferentials belong to the class of semimonotone operators. We start by studying the class of functions with upper and/or lower curvature, which are defined as follows.

Definition 4.13 (upper and lower curvature). *Let $\sigma, \ell \in \mathbb{R}$ and consider a proper lsc function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$. Then,*

   (i) (lower curvature) f is said to have lower curvature with modulus σ (denoted by $f \in \mathscr{F}_\sigma(\mathbb{R}^n)$) if $\varphi_\sigma := f - \frac{\sigma}{2}\|\cdot\|^2$ is convex.

   (ii) (upper curvature) f is said to have upper curvature with modulus ℓ (denoted by $f \in \mathscr{F}^\ell(\mathbb{R}^n)$) if f is finite-valued and $\varphi^\ell := \frac{\ell}{2}\|\cdot\|^2 - f$ is convex.

A function with both upper and lower curvatures is denoted by $\mathscr{F}^\ell_\sigma(\mathbb{R}^n)$.

Many traditional function classes fit within this curvature framework. For instance, functions with positive (resp. negative) lower curvatures are widely known as strongly (resp. hypo) convex functions. Another notable example is that of ℓ-smooth functions, which have upper curvature ℓ and lower curvature $-\ell$ [84, Lem. 2.1].

The next proposition presents the semimonotonicity properties of the subdifferential of functions according to their curvature. Additionally, a corollary is provided for the pointwise minimum of functions with upper curvatures.

Proposition 4.14 (functions with upper and lower curvature). *Let $\sigma, \ell \in \mathbb{R}$. Then, the following hold.*

   (i) If $f \in \mathscr{F}_\sigma(\mathbb{R}^n)$, i.e., f is σ-convex, then, ∂f is $(\sigma, 0)$-semimonotone.

   (ii) If f is continuous and $f \in \mathscr{F}^\ell(\mathbb{R}^n)$ with $\ell < 0$, then, ∂f is $(0, \frac{1}{\ell})$-semimonotone.

   (iii) If f is continuous and $f \in \mathscr{F}^\ell(\mathbb{R}^n)$ with $\ell \geq 0$, then, ∂f is $(\alpha(1 - c\alpha), c \,\square\, \frac{1}{\ell})$-semimonotone for all $\alpha > \ell$ and $c > \frac{1}{\alpha - \ell}$.

If $f \in \mathscr{F}^\ell_\sigma(\mathbb{R}^n)$, then, $\sigma \leq \ell$, f is continuously differentiable and the following hold.

   (iv) If $\ell + \sigma > 0$, then ∇f is $(\sigma \,\square\, \ell, \frac{1}{\ell + \sigma})$-semimonotone.

   (v) If $\ell + \sigma \leq 0$, then ∇f is $(\sigma(1 - c\sigma), c \,\square\, \frac{1}{\ell - \sigma})$-semimonotone for any constant c such that $1 + c(\ell - \sigma) > 0$.

Proof. See Appendix B.3. □

Corollary 4.15 (pointwise minimum). *Let f be the pointwise min of finite-valued and continuous functions $f_i \in \mathscr{F}^{\ell_{f_i}}(\mathbb{R}^n)$ where $\ell_{f_i} \in \mathbb{R}$, i.e., $f(x) = \min\{f_1(x), \ldots, f_N(x)\}$. Let $\bar{\ell} = \max_i \ell_{f_i}$. Then, the following hold.*

   (i) If $\bar{\ell} < 0$, then ∂f is $\left(0, \frac{1}{\bar{\ell}}\right)$-semimonotone.

   (ii) If $\bar{\ell} \geq 0$, then ∂f is $\left(\alpha(1 - c\alpha), c \,\square\, \frac{1}{\bar{\ell}}\right)$-semimonotone for all $\alpha > \bar{\ell}$ and $c > \frac{1}{\alpha - \bar{\ell}}$.

Proof. See Appendix B.3. □

One of the most fundamental operations for functions is the infimal convolution, which is the subject of the upcoming proposition. To obtain this result, we first show that the subdifferential of the infimal convolution of two semimonotone operators belongs to the parallel sum of the subdifferentials of these two functions (see Proposition 4.16(i)), and then we use Proposition 4.7(ii).

Proposition 4.16 (infimal convolution). *For $i = 1, 2$, let $f_i : \mathbb{R}^n \to \overline{\mathbb{R}}$ be proper lsc functions whose subdifferentials are (μ_i, ρ_i)-semimonotone with $\mu_i, \rho_i \in \mathbb{R}$ such that $\mu_1 + \mu_2 > 0$ or $\mu_1 = \mu_2 = 0$. Denote their infimal convolution by*

$$\varphi : s \mapsto \min_{w \in \mathbb{R}^n}\{f_1(w) + f_2(s - w)\}.$$

Suppose that for each $r_1 \geq 0$ and $\bar{s} \in \mathbb{R}^n$ there exists $r_2 > 0$ such that the set

$$\left\{(w_1, w_2) \in \mathbb{R}^n \;\middle|\; \textstyle\sum_{i=1}^2 f_i(w_i) \leq r_1, \; \left\|\sum_{i=1}^2 w_i - \bar{s}\right\| \leq r_2\right\} \tag{4.4}$$

is bounded. Then, the following hold.

(i) $\partial\varphi(\bar{s}) \subseteq (\partial f_1 \,\square\, \partial f_2)(\bar{s})$.

(ii) $\partial\varphi$ *is* $(\mu_1 \,\square\, \mu_2, \rho_1 + \rho_2)$*-semimonotone.*

Proof. See Appendix B.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

As a final example, we show that the associated saddle operator of a minimax optimization problem is semimonotone under the so-called interaction dominance condition [40].

Proposition 4.17 (saddle operator). *Suppose that* $\varphi : \mathbb{R}^n \times \mathbb{R}^m \to \overline{\mathbb{R}}$ *is twice continuously differentiable and* σ*-convex-concave for some* $\sigma \in \mathbb{R}$*, i.e.,* $\nabla^2_{xx}\varphi(z) \geq \sigma\mathrm{I}$ *and* $-\nabla^2_{yy}\varphi(z) \geq \sigma\mathrm{I}$ *for all* $z \in \mathbb{R}^{n+m}$*. Moreover, the* α*-interaction dominance condition [40, Def. 1] holds for some* $\alpha > -\frac{1}{\delta}$ *and* $\delta \in (0, 1/[-\sigma]_+)$*, i.e., for all* $z \in \mathbb{R}^{n+m}$

$$\nabla^2_{xx}\varphi(z) + \nabla^2_{xy}\varphi(z)\left(\tfrac{1}{\delta}\mathrm{I} - \nabla^2_{yy}\varphi(z)\right)^{-1}\nabla^2_{yx}\varphi(z) \geq \alpha\mathrm{I},$$

$$-\nabla^2_{yy}\varphi(z) + \nabla^2_{yx}\varphi(z)\left(\tfrac{1}{\delta}\mathrm{I} + \nabla^2_{xx}\varphi(z)\right)^{-1}\nabla^2_{xy}\varphi(z) \geq \alpha\mathrm{I}.$$

Then, the following hold for the associated saddle operator $T_\varphi(z) := (\nabla_x\varphi(z), -\nabla_y\varphi(z))$.

(i) If $\alpha < \frac{1}{\delta}$*, then* T_φ *is* $(\frac{\alpha}{1-\alpha\delta}, -\frac{\delta}{1-\alpha\delta})$*-semimonotone.*

(ii) If $\alpha \geq 0$*, then* T_φ *is* $(0, -\delta)$*-semimonotone.*

Proof. See Appendix B.3. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

5 Douglas-Rachford splitting for semimonotone operators

In practical applications, it might be difficult to determine whether the conditions of Assumption II for DRS are satisfied. To address this issue, we provide a set of sufficient conditions for the convergence of DRS involving the sum of two semimonotone operators. More specifically, we work under the following assumptions on the underlying operators A and B from problem (P-I).

Assumption III. *In problem* (P-I)*, there exist parameters* $\mu_A, \rho_A, \mu_B, \rho_B \in \mathbb{R}$ *and a nonempty set* $\mathcal{S}^\star \subseteq \mathrm{zer}\, T_{\mathrm{PD}}$ *such that for every* $(x^\star, y^\star) \in \mathcal{S}^\star$*, the following hold.*

A1 *It holds that either* $\mu_A + \mu_B > 0$ *or* $\mu_A = \mu_B = 0$*, that either* $\rho_A + \rho_B > 0$ *or* $\rho_A = \rho_B = 0$ *and that* $[\mu_A \,\square\, \mu_B]_-[\rho_A \,\square\, \rho_B]_- < \frac{1}{4}$.

A2 *Operator* A *is* (μ_A, ρ_A)*-semimonotone at* $(x^\star, -y^\star) \in \mathrm{gph}\, A$.

A3 *Operator* B *is* (μ_B, ρ_B)*-semimonotone at* $(x^\star, y^\star) \in \mathrm{gph}\, B$.

Under these assumptions, the primal-dual operator T_{PD} has oblique weak Minty solutions as shown below in Lemma 5.1*(i)*. Additionally, whenever operators A and B are maximally semimonotone as in Assumption IV, the primal-dual operator is maximally comonotone. The proof of Lemma 5.1 is deferred to Appendix B.4.

Assumption IV. *In problem* (P-I)*,* $\mathrm{zer}(A + B)$ *is nonempty and there exist parameters* $\mu_A, \rho_A, \mu_B, \rho_B \in \mathbb{R}$ *satisfying Assumption III.*A1 *such that operator* A *is maximally* (μ_A, ρ_A)*-semimonotone and operator* B *is maximally* (μ_B, ρ_B)*-semimonotone.*

Lemma 5.1 (primal-dual operator). *Suppose that either Assumption III or Assumption IV holds and let* V *be defined as in* (3.2)*, where*

$$\beta_{\mathrm{P}} = \rho_A \,\square\, \rho_B \quad and \quad \beta_{\mathrm{D}} = \mu_A \,\square\, \mu_B. \qquad\qquad\qquad\qquad (5.1)$$

Then, the following hold.

(i) If Assumption III holds, then T_{PD} *has* V*-oblique weak Minty solutions at* $\mathcal{S}^\star$.

(ii) If Assumption IV holds, then T_{PD} *is maximally* V*-comonotone.*

Table 3: *Range of γ for DRS for semimonotone operators.*

		$\mu_A = \mu_B = 0$	$\mu_A + \mu_B > 0$	
			$\mu_A\mu_B \geq 0$	$\mu_A\mu_B < 0$
$\rho_A = \rho_B = 0$		$\gamma \in (0, +\infty)$		$\gamma \in \left(0, -\frac{1}{\mu_A \square \mu_B}\right)$
$\rho_A + \rho_B > 0$	$\rho_A\rho_B \geq 0$			$\gamma \in (0, \gamma_+)$
	$\rho_A\rho_B < 0$	$\gamma \in \left(-(\rho_A \square \rho_B), +\infty\right)$	$\gamma \in (\gamma_-, +\infty)$	$\gamma \in (\gamma_-, \gamma_+)$

We are now in a position to state the main result of this section: sufficient conditions for the convergence of DRS, through the semimonotonicity properties of the operators A and B. The proof is deferred to Appendix B.4.

Theorem 5.2 (convergence of DRS under semimonotonicity). *Suppose that*

(i) either Assumption II.A1, Assumption II.A2 and Assumption III hold,

(ii) or Assumption IV hold s .

Let γ be selected according to Table 3, where

$$\gamma_- := -\frac{2(\rho_A \square \rho_B)}{1 + \sqrt{1 - 4(\mu_A \square \mu_B)(\rho_A \square \rho_B)}} \quad and \quad \gamma_+ := -\frac{1 + \sqrt{1 - 4(\mu_A \square \mu_B)(\rho_A \square \rho_B)}}{2(\mu_A \square \mu_B)}. \tag{5.2}$$

Then, $\alpha := 1 + \frac{1}{\gamma}(\rho_A \square \rho_B) + \gamma(\mu_A \square \mu_B) > 0$. Consider a sequence $(u^k, v^k, s^k)_{k \in \mathbb{N}}$ generated by DRS starting from $s^0 \in \mathbb{R}^n$ with stepsize γ and relaxation parameters $\lambda_k \in (0, 2\alpha)$ such that $\liminf_{k \to \infty} \lambda_k(2\alpha - \lambda_k) > 0$. Then, all the claims of Theorem 3.3 hold. Furthermore, if Assumption IV holds and $\lambda_k(2\alpha - \lambda_k) \geq \kappa > 0$ uniformly for all k, then the convergence estimates from Corollary 3.4 also hold.

Theorem 5.2 provides a general framework for the convergence of DRS, unifying and extending existing convergence results on relaxed Douglas/Peaceman Rachford splitting methods. We provide several examples.

Remark 5.3 (comparison with existing theory).

(i) When A and B are maximally $(0, 0)$-semimonotone, i.e., maximally monotone, the convergence result of classic DRS is recovered, see e.g. [54, 31].

(ii) When A and B are respectively μ_A- and μ_B-monotone, we recover several existing results in literature. For instance, in the strongly monotone setting where $\mu := \mu_A = \mu_B > 0$, we recover the convergence result from [59], showing that DRS converges for $\lambda \in (0, 2 + \gamma\mu)$ instead of the traditional $\lambda \in (0, 2)$. On the other hand, when $\mu_A + \mu_B > 0$, the stepsize range provided in Table 3 matches [21, Thm. 4.5(ii)]. Analogously, we recover the results for the sum of a ρ_A- and a ρ_B-comonotone operator with $\rho_A + \rho_B > 0$ from [8, Cor. 5.5(ii)].

(iii) In the optimization setting, global convergence of DRS has recently been shown in [53, 84] for structured nonsmooth, nonconvex optimization problems where the cost is split as the sum of one Lipschitz differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ and one proper lsc function $g : \mathbb{R}^n \to \mathbb{R}$, i.e., $A = \nabla f$ and $B = \partial g$. These results rely on showing descent of the iterates on the so-called Douglas-Rachford envelope [65], for which only properties of the Lipschitz differentiable function f are exploited to determine the range of admissible stepsizes. For instance, consider the setting where f is ℓ-smooth and σ-convex, where $\ell > 0$ and $\sigma \in [-\ell, 0)$. Then, it is shown in [84] that DRS converges globally for $\lambda \in (0, 2)$ and $\gamma < \min\{\frac{\lambda - 2}{2\sigma}, \frac{1}{\ell}\} < -\frac{1}{\sigma}$ provided that $\arg\min_x f(x) + g(x)$ is nonempty.

Note that when additional semimonotonicity properties of g are given, a larger stepsize range for γ can be obtained through Theorem 5.2. To see this, first observe that in this setting A is (μ_A, ρ_A)-semimonotone for some $\mu_A < 0$ and $\rho_A \geq 0$ (see Proposition 4.14(iv)). Therefore, supposing that ∂g is (μ_B, ρ_B)-semimonotone at $(x^\star, y^\star) \in \operatorname{gph} \partial g$ where $\mu_A + \mu_B > 0$ and $\rho_B \geq 0$ it follows from Theorem 5.2 that DRS converges for $\gamma \in (0, \gamma_+)$, which is a larger stepsize range since $\gamma_+ > -\frac{1}{\sigma}$. $\quad\square$

6 Example problems

In this section, we provide several example problems for DRS that go beyond the standard monotone setting. Details of the examples presented in this section can be found in Appendix C.

We commence by applying DRS to a simple linear inclusion problem. This setting is of particular interest for two specific reasons. First, this example demonstrates the tightness of the stepsize range provided in Theorem 3.3 (see Example 6.1(ii)). Second, it highlights that our theory is able to cover cases where neither the primal nor the dual nor the primal-dual inclusion is monotone (see Example 6.1(iii)).

Example 6.1 (saddle point problem). Consider the problem of finding a stationary point of a linear inclusion, with a particular structure that emerges naturally in saddle point problems [33, §1.4.1]

$$0 \in T_P x = \begin{bmatrix} b & a \\ -a & b \end{bmatrix} x = \underbrace{\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}}_{A} x + \underbrace{\begin{bmatrix} b & 0 \\ 0 & b \end{bmatrix}}_{B} x, \tag{6.1}$$

where $a, b \in \mathbb{R} \setminus \{0\}$. In particular, any solution to the inclusion problem $0 \in T_P x$ is a minimax solution of $f(x_1, x_2) := a x_1 x_2 + \frac{b}{2}(x_1^2 - x_2^2)$ when $b > 0$ and a maximin solution when $b < 0$. For this problem, the following assertions hold.

(i) By examining the spectral radius of the algorithmic operator, it can be seen that the sequence $(u^k, v^k, s^k)_{k \in \mathbb{N}}$ generated by DRS with fixed relaxation parameter λ converges iff λ lies in the interval

$$0 < \lambda < \bar{\lambda} := 2 \frac{(a^2 \gamma + b)(1 + \gamma b)}{\gamma(a^2 + b^2)}, \tag{6.2}$$

which is nonempty iff either $b > 0$ or

$$b < 0, \ a^2 \neq b^2 \text{ and } \gamma \in \left(\min\{-\tfrac{1}{b}, -\tfrac{b}{a^2}\}, \max\{-\tfrac{1}{b}, -\tfrac{b}{a^2}\} \right).$$

A numerical experiment verifying this convergence result is provided in Figure 6.

Figure 6: *Sequences u^k generated by applying DRS on Example 6.1 for $a = 2$, $b = -1$ and stepsize $\gamma = \frac{1}{2}$.*

(ii) The results from Theorem 3.3 are tight in the sense that the entire range of relaxation parameters λ from (6.2) is covered, when in (2.1) the vector v is restricted to $\mathcal{R}(P)$ (see the remark below Definition 2.1).

(iii) The range of a and b from Example 6.1(i) for which DRS converges covers cases where neither the primal, nor the dual, nor the primal-dual problem is monotone.
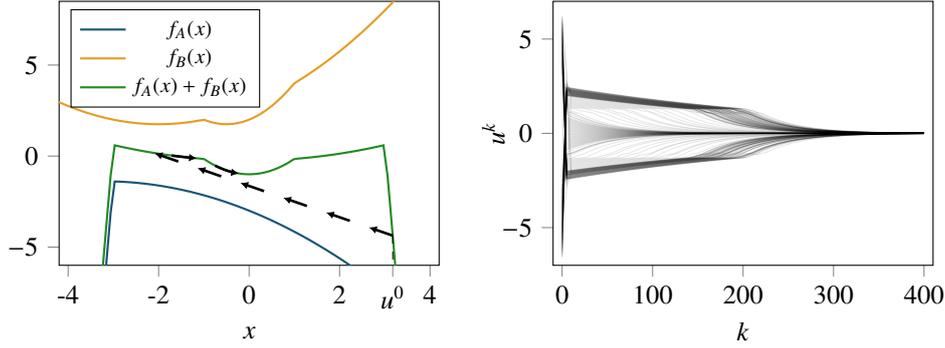
Additionally, an analysis of the tightness of Theorem 5.2 is provided in Appendix C. In this analysis, the semimonotonicity parameters $\mu_A, \mu_B, \rho_A, \rho_B$ are selected according to Proposition 4.5 such that the relaxation parameter in Theorem 5.2 is maximized. Notice that although A is obviously $(0, 0)$-semimonotone, the maximal relaxation parameter is in general attained for other values of μ_A and ρ_A. ☐

Next, we turn our attention to a nonsmooth optimization problem. This example highlights the generality of our underlying assumptions, as the convergence of DRS for this problem is not covered by any previous works on DRS (to the best of the authors' knowledge).

Example 6.2 (nonsmooth optimization). Consider the separable nonsmooth optimization problem

Figure 8: *(left) Visualization of the functions involved in Example 6.4. Additionally, the sequence $(u^k)_{k=0}^{400}$ is indicated using black arrows, which is obtained by applying DRS to $0 \in \partial f_A(x) + \partial f_B(x)$ with stepsize $\gamma = {}^{11}/_{60} \in ({}^1/_6, {}^1/_5)$ and relaxation $\lambda = {}^9/_{10}(2 - {}^4/_{15\gamma} - {}^{12}\gamma/_5) \approx 0.095$. When evaluating the (multi-valued) resolvents $J_{\gamma A}$ and $J_{\gamma B}$, each element is sampled with equal probability. (right) Visualization of the sequence $(u^k)_{k=0}^{400}$ for 200 such experiments, of which the initializations s^0 are evenly spaced within the interval $[-5, 5]$.*

$$\underset{x \in \mathbb{R}}{\text{minimize}} \quad f_A(x) + f_B(x)$$

which is visualized in Figure 7, where

$$f_A(x) := \begin{cases} 3x^2 - 3x + 8, & \text{if } x \in [-1, 1], \\ -\frac{1}{2}x^2 - 3x + \frac{23}{2}, & \text{otherwise,} \end{cases}$$

$$f_B(x) := \begin{cases} 2x^2 - x, & \text{if } x < -1, \\ x^2 + 3x + 5, & \text{if } x \in [-1, 2], \\ x^2 + 15x - 19, & \text{if } x > 2. \end{cases}$$

The following hold for the inclusion $0 \in T_P x = \partial f_A(x) + \partial f_B(x)$.

Figure 7: *Visualization of the (nonsmooth) functions involved in Example 6.2.*

(i) Assumption II.A1 and Assumption II.A2 hold when $\gamma \neq 1$.

(ii) Assumption III holds for $\mathcal{S}^\star = \{(0, 3)\}$ and $\mu_A = -1.2$, $\rho_A = 0.2$, $\mu_B = 1.6$ and $\rho_B = 0.1$.

(iii) The claims from Theorem 3.3 hold for the sequence $(u^k, v^k, s^k)_{k \in \mathbb{N}}$ generated by DRS with stepsize $\gamma \in (0, 0.2614)$ and relaxation parameter $\lambda \in (0, 2 + {}^2/_{15\gamma} - {}^{48}\gamma/_5)$. $\quad\square$

Remark 6.3 (comparison with previous works).

(i) Since both f_A and f_B are nonsmooth and f_A is nonconvex, convergence results for DRS in the optimization setting such as [53, 84] cannot be applied.

(ii) Since $\rho_A > 0$ and $\rho_B > 0$, it holds by definition that ∂f_A is μ_A-monotone at $(x^\star, -y^\star)$ and ∂f_B is μ_B-monotone at $(x^\star, y^\star)$. However, as the monotonicity of A and B does not hold globally, the results from [21] are not applicable here. $\quad\square$

Next, consider the following example, where we apply DRS to the subdifferential of a separable nonsmooth function. For this example, we show that our analysis of DRS covers its convergence behavior towards the stationary points, and provide numerical results in Figure 8. Notice that in the iterates are attracted by the local minimum, while in contrast e.g. descent methods would quickly diverge.

Example 6.4 (stationary point). Let $\mu_A = -0.3$, $\rho_A = -0.1$, $\mu_B = 0.4$ and $\rho_B = 0.4$. Consider the problem of finding a stationary point of $f_A(x) + f_B(x)$, visualized in Figure 8, where

$$f_A(x) := \begin{cases} \frac{1 - \sqrt{1 - 4\mu_A \rho_A}}{4\rho_A} x^2 - x - 3, & \text{if } x \in [-3, 3], \\ \frac{1 + \sqrt{1 - 4\mu_A \rho_A}}{4\rho_A} x^2 - x - \frac{9\sqrt{1 - 4\mu_A \rho_A}}{2\rho_A} - 3, & \text{otherwise,} \end{cases}$$

26

$$f_B(x) := \begin{cases} x^2 + x + 2, & \text{if } x \in [-1, 1], \\ \frac{1}{4}x^2 + x + \frac{11}{4}, & \text{otherwise.} \end{cases}$$

The following hold for the inclusion $0 \in T_P x = \partial f_A(x) + \partial f_B(x)$.

(i) Assumption II.A1 and Assumption II.A2 hold for $\gamma \in (0, \frac{-2\rho_A}{1+\sqrt{1-4\mu_A\rho_A}}) \cup (\frac{-2\rho_A}{1+\sqrt{1-4\mu_A\rho_A}}, 1/5)$.

(ii) Assumption III holds for $\mathcal{S}^\star = \{(0, -1)\}$ and the given $\mu_A, \rho_A, \mu_B, \rho_B$.

(iii) The claims from Theorem 3.3 hold for the sequence $(u^k, v^k, s^k)_{k\in\mathbb{N}}$ generated by DRS with stepsize $\gamma \in (1/6, 1/5)$ and fixed relaxation parameter $\lambda \in (0, 2 - 4/15\gamma - 12\gamma/5)$ (see Figure 8 for numerical results). □

As a final example, consider the following quadratic programming (QP) problem, with box and linear equality constraints. When Q is positive semidefinite, which is the convex setting, convergence of DRS for Example 6.5 follows directly by monotonicity of A and B. The following example demonstrates that our results allow to go beyond the monotone setting, leveraging the calculus rules from Section 4. Note that in this example both f and g are nonsmooth, and thus convergence results for DRS such as [53, 84] cannot be applied here.

Example 6.5 (constrained QP). Consider the following quadratic program

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \tfrac{1}{2}x^\top Q x + q^\top x \\ \text{subject to} \quad & Hx = h, \\ & l \le x \le u, \end{aligned} \tag{6.3}$$

where $Q \in \mathbb{S}^n$, $q \in \mathbb{R}^n$, $H \in \mathbb{R}^{m\times n}$ has full row rank, $h \in \mathbb{R}^m$ and scalars $l, u \in \mathbb{R}$ satisfying $l < u$. This problem can be equivalently formulated as $\text{minimize}_x f(x) + g(x)$ where

$$f(x) := \tfrac{1}{2}x^\top Q x + q^\top x + \delta_{\{s \mid Hs=h\}}(x) \quad \text{and} \quad g(x) := \delta_{\{s \mid l\le s\le u\}}(x),$$

and the associated first-order optimality condition is given by $0 \in A(x) + B(x) := \partial f(x) + \partial g(x)$, By splitting the quadratic program (6.3) into f and g, the resulting DRS iterations can be computed efficiently in closed form. In particular, f corresponds to an equality constrained QP problem, of which the prox can be solved efficiently through its corresponding KKT system [62, §16.1], and computing the resolvent of ∂g corresponds to projecting onto a box.

For instance, let $Q = \text{diag}(-1, \frac{1}{2})$, $q = [-4 \; -1]^\top$, $H = [1 \; 1]$, $h = 1$, $l = 0$ and $u = 1$. Then, the unique minimizer of (6.3) is given by $x^\star = [1 \; 0]^\top$ and the following assertions hold.

(i) Operator A is $(-1, 0)$-semimonotone.

(ii) Operator B is $(2, 0)$-semimonotone at $(x^\star, y^\star) = ([1 \; 0]^\top, [2 \; -2]^\top) \in \text{gph } B$.

(iii) The sequence $(u^k, 1/\gamma(u^k - s^k))_{k\in\mathbb{N}}$ generated by DRS with fixed relaxation parameter λ converges to $\text{zer } T_{\text{PD}}$ for $\gamma \in (0, 1/2)$ and $\lambda \in (0, 2 - 4\gamma)$.

Proof. By Propositions 4.4*(ii)* and 4.5, it follows that the mapping $x \mapsto Qx + q$ is $(-1, 0)$-semimonotone. The claimed semimonotonicity of A then follows by observing that the normal cone $N_{\{s \mid Hs=h\}}$ is monotone [9, Ex. 20.26] and applying Proposition 4.7*(i)*. Finally, the claimed semimonotonicity of B follows from Proposition A.7 and the convergence result for DRS follows from Theorem 5.2. □

7 Conclusion

This paper presented a comprehensive study of the (relaxed) preconditioned proximal point algorithm (PPPA) and the (relaxed) Douglas-Rachford splitting (DRS) method for a class of nonmonotone problems that satisfy an oblique weak Minty condition. To make the achieved results for DRS readily applicable in practice, the class of semimonotone operators was introduced, and sufficient conditions were developed for the convergence of DRS for the sum of two semimonotone operators. Various relevant example problems were provided thoughout the paper, showing tightness of the achieved convergence results.

Future research directions include investigating inexact variants of PPPA, analyzing other splitting methods in nonmonotone settings, and studying applications in noncooperative game theory and nonconvex distributed optimization.

A Auxiliary lemmas

Lemma A.1. *Let $P \in \mathbb{S}_+^n$ be a symmetric positive semidefinite matrix, U be an orthonormal basis for $\mathcal{R}(P)$ and define $Q := P + \Pi_{\mathcal{N}(P)}$. Then, the following properties hold.*

(i) The projection onto the range of P is given by $\Pi_{\mathcal{R}(P)} = UU^\top$.

(ii) $P = P\Pi_{\mathcal{R}(P)} = \Pi_{\mathcal{R}(P)} P = Q\Pi_{\mathcal{R}(P)}$.

(iii) $Q > 0$.

(iv) $U^\top PU = U^\top QU > 0$.

Proof. Statement A.1*(i)* follows from [57, §5.13] and A.1*(ii)* follows directly by definition of P and Q.

♠ A.1*(iii)*: Let $x \in \mathbb{R}^n \setminus \{0\}$ and consider decomposing it as $x = x_1 + x_2$ where $x_1 \in \mathcal{R}(P)$ and $x_2 \in \mathcal{N}(P)$. Let $U_1\Lambda_1 U_1^\top$ be a compact eigendecomposition of P, so that $U_1 U_1^\top = \Pi_{\mathcal{R}(P)}$ and $p_{\min} := \lambda_{\min}(\Lambda_1) > 0$. Then,

$$\langle x, Qx \rangle = \langle x, Px \rangle + \langle x, \Pi_{\mathcal{N}(P)} x \rangle = \langle x_1, U_1\Lambda_1 U_1^\top x_1 \rangle + \|x_2\|^2$$
$$\geq p_{\min}\langle x_1, \Pi_{\mathcal{R}(P)} x_1 \rangle + \|x_2\|^2 = p_{\min}\|x_1\|^2 + \|x_2\|^2 \geq \min\{1, p_{\min}\}\|x\|^2 > 0.$$

♠ A.1*(iv)*: The equivalence $U^\top PU = U^\top QU$ follows from Lemmas A.1*(i)* and A.1*(ii)* using $U^\top U = \mathrm{I}$. Positive definiteness follows from Lemma A.1*(iii)* and the fact that U has full column rank [11, Prop. 8.1.2(xiii)]. □

Proposition A.2 (resolvent of a comonotone operator)**.** *Let $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be a V-comonotone operator and suppose that P is selected according to Assumption I.A4. Then, the preconditioned resolvent $(P + T)^{-1} \circ P$ has full domain if and only if T is maximal.*

Proof. Since T is V-comonotone, it holds by definition that $T^{-1} - V$ is monotone and thus also that $\tilde{T} := U^\top T^{-1} U - U^\top VU$ is monotone. Furthermore, it holds that $(U^\top PU)^{-1} + U^\top VU > 0$ by Assumption I.A4 and Remark 2.2. Denote $\hat{T} := U^\top T^{-1}U + (U^\top PU)^{-1}$. Owing to $(U^\top PU)^{-1} + U^\top VU > 0$ and Minty's theorem [9, Thm. 21.1], the operator

$$((U^\top PU)^{-1} + U^\top VU)^{-1}\hat{T} = ((U^\top PU)^{-1} + U^\top VU)^{-1}\tilde{T} + \mathrm{id}$$

has full range iff $((U^\top PU)^{-1} + U^\top VU)^{-1}\tilde{T}$ is maximally monotone. Since $(U^\top PU)^{-1} + U^\top VU > 0$, this implies that $\hat{T}$ has full range iff $\tilde{T}$ is maximally monotone, i.e., iff T is maximally V-comonotone.

Therefore, it only remains to show that $\hat{T}$ has full range iff $\mathrm{dom}\,(P + T)^{-1} \circ P = \mathbb{R}^n$. First suppose that $\hat{T}$ has full range. Let m denote the dimension of the range of P and fix $\hat{y} \in \mathbb{R}^m$. Then, by the full range assumption for $\hat{T}$, there exists some $\hat{x} \in \mathbb{R}^m$ such that $\hat{y} \in \hat{T}\hat{x}$. Letting $\hat{z} = \hat{y} - (U^\top PU)^{-1}\hat{x}$, this implies that

$$\hat{y} \in U^\top T^{-1}U\hat{x} + (U^\top PU)^{-1}\hat{x} \iff \hat{z} \in U^\top T^{-1}U(U^\top PU(\hat{y} - \hat{z})) \tag{A.1}$$

Multiplying by U, this implies that $U\hat{z} \in \Pi_{\mathcal{R}(P)} T^{-1}(PU(\hat{y} - \hat{z}))$. Defining $x' = U\hat{x}$, $y' = U\hat{y}$ and $z' = U\hat{z}$, this corresponds to $z' \in \Pi_{\mathcal{R}(P)} T^{-1}(P(y' - z'))$. Consequently, $T^{-1}(P(y' - z'))$ is nonempty and there exists a $z'' \in \mathcal{N}(P) \subseteq \mathbb{R}^n$ so that $z'' \in \Pi_{\mathcal{N}(P)} T^{-1}(P(y' - z'))$. Summing both inclusions, it holds that

$$z' + z'' \in T^{-1}(P(y' - z')) \iff P(y' - z') \in T(z' + z'') \iff z' + z'' \in (P + T)^{-1}Py' \tag{A.2}$$

Since the choice of y' is arbitrary in $\mathcal{R}(P)$, $(P + T)^{-1}P$ has full domain.

The converse follows similarly. Assume that $\mathrm{dom}\,(P + T)^{-1}P = \mathbb{R}^n$. Then, for any fixed $\hat{y} \in \mathbb{R}^m$ and $y' = U\hat{y}$ there exists some $z \in \mathbb{R}^n$ such that $z \in (T + P)^{-1}Py$. Without loss of generality, decompose $z := U\hat{z} + z''$, where $\hat{z} \in \mathbb{R}^m$ and $z'' \in \mathcal{N}(P)$. Then, by (A.2) it follows that $U\hat{z} + z'' \in T^{-1}(PU(\hat{y} - \hat{z}))$. Multiplying by $U^\top$, this implies that $\hat{z} \in U^\top T^{-1}U(U^\top PU(\hat{y} - \hat{z}))$. Defining $\hat{x} = U^\top PU(\hat{y} - \hat{z})$ and using (A.1), we obtain $\hat{y} \in \hat{T}\hat{x}$, thus establishing that $\hat{T}$ has full range. □

Fact A.3 (Solution of quadratic inequality)**.** *Let $a, b, c \in \mathbb{R}$ and suppose that $b > -2\sqrt{ac}$ when $ac \geq 0$. Then, the following hold.*

Table 4: *Ranges of γ for which the resolvents $J_{\gamma A}$ and $J_{\gamma B}$ have full domain.*

		$\mu_A = \mu_B = 0$	$\mu_A + \mu_B > 0$	
			$\mu_A \mu_B \geq 0$	$\mu_A \mu_B < 0$
$\rho_A = \rho_B = 0$		$\gamma \in (0, +\infty)$		$\gamma \in \left(0, -\frac{1}{\min\{\mu_A, \mu_B\}}\right)$
$\rho_A + \rho_B > 0$	$\rho_A \rho_B \geq 0$			$\gamma \in (0, \bar{\gamma}_+)$
	$\rho_A \rho_B < 0$	$\gamma \in (-\min\{\rho_A, \rho_B\}, +\infty)$	$\gamma \in (\bar{\gamma}_-, +\infty)$	$\gamma \in (\bar{\gamma}_-, \bar{\gamma}_+)$

(i) *There exists a $\gamma > 0$ satisfying $a\gamma^2 + b\gamma + c > 0$ if and only if it holds that*

$$b > 2\sqrt{ac} \quad when \quad a < 0 \text{ and } c < 0. \tag{A.3}$$

(ii) *If (A.3) holds, then $\gamma > 0$ satisfies $a\gamma^2 + b\gamma + c > 0$ if and only if $\gamma \in \left(\frac{2[-c]_+}{b + \sqrt{b^2 - 4ac}}, \frac{b + \sqrt{b^2 - 4ac}}{2[-a]_+}\right)$, which is nonempty.*

Fact A.4. *Let $a, b \in \mathbb{R}$ such that $a + b > 0$. Then, $a \square b < \min\{a, b\}$.*

Lemma A.5 (implications of Assumption II). *If Assumption II holds, then Assumptions I.A1 to I.A3 hold for operator T_{PD} from (PD-I) and preconditioner P as in (3.1).*

Proof. First, outer semicontinuity of T_{PD} follows from that of A and B [77, Thm. 5.7(a)], showing I.A1. Second, I.A2 holds since $J_{\gamma A}$ and $J_{\gamma B}$ having full domain implies that the preconditioned resolvent $(P + T_{\mathrm{PD}})^{-1}P$ has full domain, owing to [77, Lem. 12.14]. Finally, statement I.A3 is immediate, completing the proof. □

Lemma A.6 (implications of Assumption IV). *Suppose that Assumption IV holds and that γ is selected according to Table 3, where γ_- and γ_+ are defined as in (5.2). Then, Assumptions II.A1 and II.A2 hold and the resolvents $J_{\gamma A}$ and $J_{\gamma B}$ are Lipschitz continuous.*

Proof. First, observe that $[\mu_A \square \mu_B]_- [\rho_A \square \rho_B]_- < \frac{1}{4}$, which by Fact A.4 and Proposition 4.3(ii) implies that $[\mu_A]_- [\rho_A]_- < \frac{1}{4}$ and $[\mu_B]_- [\rho_B]_- < \frac{1}{4}$. Therefore, owing to Proposition 4.11, it follows that A and B are outer semicontinuous, showing that Assumption II.A1 holds. Furthermore, owing to Proposition 4.12(i), it follows that the resolvents $J_{\gamma A}$ and $J_{\gamma B}$ are Lipschitz continuous and have full domain if γ complies with Table 4, where

$$\bar{\gamma}_- = \begin{cases} \frac{-2\rho_A}{1 + \sqrt{1 - 4\mu_A \rho_A}}, & \text{if } \rho_A < 0, \\ \frac{-2\rho_B}{1 + \sqrt{1 - 4\mu_B \rho_B}}, & \text{otherwise,} \end{cases} \quad \text{and} \quad \bar{\gamma}_+ = \begin{cases} -\frac{1 + \sqrt{1 - 4\mu_A \rho_A}}{2\mu_A}, & \text{if } \mu_A < 0, \\ -\frac{1 + \sqrt{1 - 4\mu_B \rho_B}}{2\mu_B}, & \text{otherwise.} \end{cases} \tag{A.4}$$

As γ is selected according to Table 3, it simply remains to prove that these ranges are a subset of the ranges for γ provided in Table 4, which can be verified directy using Fact A.4. □

Proposition A.7 (normal cone of a box). *The normal cone operator $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of the n-dimensional box $C := \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i, \ i = 1, \ldots, n\}$ is $\left(\min\left\{\frac{|\tilde{v}_1|}{u_1 - l_1}, \ldots, \frac{|\tilde{v}_n|}{u_n - l_n}\right\}, 0\right)$-semimonotone at $(\tilde{x}, \tilde{v}) \in \mathrm{gph}\, N_C$.*

Proof. By monotonicity of the normal cone N_C [9, Ex. 20.26], it holds that

$$\langle \tilde{v}, \tilde{x} - x \rangle = \sum_{i=1}^n |\tilde{v}_i||\tilde{x}_i - x_i| \geq \sum_{i=1}^n \frac{|\tilde{v}_i|}{u_i - l_i}|\tilde{x}_i - x_i|^2 \geq \min\left\{\frac{|\tilde{v}_1|}{u_1 - l_1}, \ldots, \frac{|\tilde{v}_n|}{u_n - l_n}\right\}\|\tilde{x} - x\|^2,$$

where the first inequality holds since $|\tilde{x}_i - x_i| \leq u_i - \ell_i$ for all $i \in \{1, \ldots, n\}$. □

B Omitted proofs

B.1 Douglas-Rachford splitting

Proof of Lemma 3.1 *(equivalence of DRS and PPPA).* The update rule (PPPA) applied to (PD-I) with preconditioner (3.1) can be equivalently written as

$$\bar{x}^k \in J_{\gamma A}(x^k - \gamma y^k) \tag{B.1a}$$

$$\bar{y}^k \in (\gamma \mathrm{id} + B^{-1})^{-1}(2\bar{x}^k - x^k + \gamma y^k) \Leftrightarrow 2\bar{x}^k - x^k + \gamma y^k - \gamma \bar{y}^k \in J_{\gamma B}(2\bar{x}^k - x^k + \gamma y^k) \tag{B.1b}$$

$$x^{k+1} = x^k + \lambda_k(\bar{x}^k - x^k) \tag{B.1c}$$

$$y^{k+1} = y^k + \lambda_k(\bar{y}^k - y^k), \tag{B.1d}$$

where [77, Lem. 12.14] was used in (B.1b). We proceed by induction. The base case $k = 0$ is vacuously satisfied through initialization $s^0 = x^0 - \gamma y^0$, and by using (B.1a) and (B.1b), noting that for any $\bar{x}^0, \bar{y}^0$ there exists $u^0 = \bar{x}^0 \in J_{\gamma A}(s^0)$, $v^0 = 2u^0 - s^0 - \gamma \bar{y}^0 \in J_{\gamma B}(2u^0 - s^0)$ (and vice versa). Suppose now that the claim holds for some $k \geq 0$. In particular, $s^k = x^k - \gamma y^k$, $u^k = \bar{x}^k$ and $v^k = 2u^k - s^k - \gamma \bar{y}^k$. According to update rule (DRS), it holds that

$$s^{k+1} = s^k + \lambda_k(v^k - u^k) = s^k + \lambda_k(\bar{x}^k - s^k - \gamma \bar{y}^k) = x^k - \gamma y^k + \lambda_k(\bar{x}^k - \gamma \bar{y}^k - x^k + \gamma y^k) = x^{k+1} - \gamma y^{k+1},$$

where the induction hypothesis was used in the first two equalities, and (B.1c) and (B.1d) were used in the last equality. Having established $s^{k+1} = x^{k+1} - \gamma y^{k+1}$, the claim for the correspondences between $\bar{x}^{k+1}, \bar{y}^{k+1}$ and u^{k+1}, v^{k+1} follow from (B.1a) and (B.1b) as argued in the base case. $\square$

Proof of Lemma 3.2 *(weak Minty for primal and dual operator).* Let $(x_A, y_A) \in \mathrm{gph}\, A$ and $(x_B, y_B) \in \mathrm{gph}\, B$. Then, by definition of the primal-dual operator and using that $(y_B, x_B) \in \mathrm{gph}\, B^{-1}$ it holds that $(z, v) := \big((x_A, y_B), (y_A + y_B, x_B - x_A)\big) \in \mathrm{gph}\, T_{\mathrm{PD}}$. Therefore, by Assumption II.A3 it holds for all $z^\star = (x^\star, y^\star) \in \mathcal{S}^\star$ that $\langle v, z - z^\star \rangle \geq \mathrm{q}_V(v)$, i.e.,

$$\big\langle y_A + y_B, x_A - x^\star \big\rangle + \big\langle x_B - x_A, y_B - y^\star \big\rangle \geq \beta_{\mathrm{P}}\|y_A + y_B\|^2 + \beta_{\mathrm{D}}\|x_B - x_A\|^2. \tag{B.2}$$

Let $x_A = x_B =: x_P$. Then, from (B.2) it follows that $\langle y_A + y_B, x_P - x^\star \rangle \geq \beta_{\mathrm{P}}\|y_A + y_B\|^2$ for all $x^\star \in \mathcal{S}_P^\star$. Since $(x_P, y_A + y_B) \in \mathrm{gph}\, T_{\mathrm{P}}$ it is of immediate verification that T_{P} has β_{P}-weak Minty solutions at $\mathcal{S}_P^\star$.

Similarly, let $-y_A = y_B =: y_D$. Then, due to (B.2) $\langle x_B - x_A, y_D - y^\star \rangle \geq \beta_{\mathrm{D}}\|x_B - x_A\|^2$ for all $y^\star \in \mathcal{S}_D^\star$ and the claim for T_{D} follows analogously from the fact that $(y_D, x_B - x_A) \in \mathrm{gph}\, T_{\mathrm{D}}$, completing the proof. $\square$

Proof of Theorem 3.6 *(linear convergence of DRS for piecewise polyhedral mappings).*
Since operator B is piecewise polyhedral so is its inverse, and in turn so is the mapping $\hat{T} = A \times B^{-1}$. Therefore, T_{PD} being the sum of $\hat{T}$ and a linear mapping is also piecewise polyhedral. Since the inverse of a piecewise polyhedral mapping is piecewise polyhedral, it follows from [25, 3H.1 and 3H.3] that T_{PD} is metrically subregular at any z for any v with $(z, v) \in \mathrm{gph}\, T_{\mathrm{PD}}$. The claimed rates for DRS are an immediate consequence of Theorem 2.13, owing to the relations $U_1^\top z^k = s^k$ and (3.5). $\square$

B.2 Semimonotone operators

Proof of Proposition 4.3 *(existence of semimonotone operators).* Consider $(x, y), (x', y') \in \mathrm{gph}\, A$ and let $\epsilon > 0$. Then, by the Fenchel-Young inequality, it holds that

$$-\epsilon\|x - x'\|^2 - \epsilon^{-1}\|y - y'\|^2 \leq 2\langle x - x', y - y' \rangle \leq \epsilon\|x - x'\|^2 + \epsilon^{-1}\|y - y'\|^2, \tag{B.3}$$

♠ 4.3(i): The first inequality of (B.3) implies that (4.1) is always satisfied for $\mu \leq -\frac{\epsilon}{2}$ and $\rho \leq -\frac{1}{2\epsilon}$, i.e., for $[\mu]_-[\rho]_- \geq \frac{1}{4}$, which completes the proof.

♠ 4.3*(ii)*: The second inequality of (B.3) implies that (4.1) can never be satisfied when $\mu > \frac{\epsilon}{2}$ and $\rho > \frac{1}{2\epsilon}$, i.e., when $[\mu]_+[\rho]_+ > \frac{1}{4}$.

♠ 4.3*(iii)*: The second inequality of (B.3) implies that (4.1) can only be satisfied with equality when $\mu = \frac{\epsilon}{2} > 0$ and $\rho = \frac{1}{2\epsilon} > 0$, i.e.,

$$\langle x - x', y - y' \rangle = \mu \|x - x'\| + \frac{1}{4\mu} \|y - y'\|^2, \qquad \text{for all } (x, y), (x', y') \in \text{gph } A,$$

which is equivalent to

$$\mu \left\| x - x' - \frac{1}{2\mu}(y - y') \right\|^2 = 0, \qquad \text{for all } (x, y), (x', y') \in \text{gph } A.$$

By construction, this may only hold if $\text{gph } A \subseteq \text{gph } T$ for some $c \in \mathbb{R}^n$. Since $\text{gph } T$ cannot be enlarged by another point $(\hat{x}, \hat{y})$ while maintaining (μ, ρ)-semimonotonicity, T is maximally (μ, ρ)-semimonotone, and the proof is complete. $\qquad \square$

Proof of Corollary 4.6 *(multiple of identity).* From Proposition 4.5 and using that $A = \alpha I_n$, it follows that A is (μ, ρ)-semimonotone if and only if $\alpha \geq \mu + \rho\alpha^2$, which is trivially satisfied when $(\mu, \rho) = (\alpha(1 - c\alpha), c)$ for any $c \in \mathbb{R}$. Now, let $(\hat{x}, \hat{y}) \in \mathbb{R}^n \times \mathbb{R}^n$ and suppose that

$$\langle x - \hat{x}, \alpha x - \hat{y} \rangle \geq \alpha(1 - c\alpha)\|x - \hat{x}\|^2 + c\|\alpha x - \hat{y}\|^2, \qquad \forall x \in \mathbb{R}^n. \tag{B.4}$$

To show when A is maximally $(\alpha(1 - c\alpha), c)$-semimonotone, we must show when it is guaranteed that $(\hat{x}, \hat{y}) \in \text{gph } A$, i.e., that $\hat{y} = \alpha\hat{x}$. Through elementary algebra, it can be shown that (B.4) is equivalent to

$$0 \geq (1 - 2c\alpha)\langle x, \hat{y} - \alpha\hat{x} \rangle + \alpha(1 - c\alpha)\|\hat{x}\|^2 - \langle \hat{x}, \hat{y} \rangle + c\|\hat{y}\|^2, \qquad \forall x \in \mathbb{R}^n. \tag{B.5}$$

Due to the dependence of the first term on x, necessarily either $\hat{y} = \alpha\hat{x}$ or $c = \frac{1}{2\alpha}$. Plugging in $\hat{y} = \alpha\hat{x}$ into (B.5), we get $0 \geq 0$ (as expected). Plugging in $c = \frac{1}{2\alpha}$ into (B.5), we get

$$0 \geq \frac{1}{2}\alpha\|\hat{x}\|^2 - \langle \hat{x}, \hat{y} \rangle + \frac{1}{2\alpha}\|\hat{y}\|^2 = \frac{1}{2\alpha}\|\alpha\hat{x} - \hat{y}\|^2, \qquad \forall x \in \mathbb{R}^n.$$

Therefore, (B.5) holds iff either (i) $\hat{y} = \alpha\hat{x}$ or (ii) $c = \frac{1}{2\alpha}$ and $\alpha < 0$, completing the proof. $\qquad \square$

Proof of Proposition 4.7 *(sum and parallel sum).*

♠ 4.7*(i)*: First, consider the case where A and B are only semimonotone at a single point. Let $(x, y_A) \in \text{gph } A$ and $(x, y_B) \in \text{gph } B$. Then, it holds that

$$\langle x - x', y_A + y_B - (y'_A + y'_B) \rangle = \langle x - x', y_A - y'_A \rangle + \langle x - x', y_B - y'_B \rangle$$
$$\geq (\mu_A + \mu_B)\|x - x'\|^2 + \rho_A\|y_A - y'_A\|^2 + \rho_B\|y_B - y'_B\|^2.$$

Consequently, by definition $A + B$ is $(\mu_A + \mu_B, \rho)$-semimonotone at $(x', y'_A + y'_B)$ if and only if

$$\rho\|y_A + y_B - (y'_A + y'_B)\|^2 = \rho\|y_A - y'_A\|^2 + 2\rho\langle y_A - y'_A, y_B - y'_B \rangle + \rho\|y_B - y'_B\|^2 \leq \rho_A\|y_A - y'_A\|^2 + \rho_B\|y_B - y'_B\|^2,$$

which is equivalent to the LMI $\left(\begin{smallmatrix} \rho_A - \rho & -|\rho| \\ -|\rho| & \rho_B - \rho \end{smallmatrix} \right) \succeq 0$. This is satisfied if and only if the determinant and trace are nonnegative, i.e.,

$$\rho_A\rho_B - \rho(\rho_A + \rho_B) \geq 0 \text{ and } \rho_A + \rho_B - 2\rho \geq 0 \quad \Leftrightarrow \quad \rho \leq \begin{cases} \frac{\rho_A\rho_B}{\rho_A + \rho_B}, & \text{if } \rho_A + \rho_B > 0, \\ 0, & \text{if } \rho_A = \rho_B = 0. \end{cases} \tag{B.6}$$

Since by definition the largest possible value for ρ satisfying (B.6) is given by $\rho_A \square \rho_B$ (if either $\rho_A + \rho_B > 0$ or $\rho_A = \rho_B = 0$), the first part of the proof is completed.

If A and B are semimonotone then by definition A is (μ_A, ρ_A)-semimonotone at all $(x'_A, y'_A) \in \text{gph } A$ and B is (μ_B, ρ_B)-semimonotone at all $(x'_B, y'_B) \in \text{gph } B$. Hence, using the previous result we know that $A + B$ is $(\mu_A + \mu_B, \rho_A \square \rho_B)$-semimonotone at all points in the set

$$\{(x', v'_A + v'_B) \mid x' \in \text{dom } A \cap \text{dom } B, \ v'_A \in Ax', \ v'_B \in Bx'\}.$$

Since this set is equal to $\text{gph}(A + B)$, the proof is completed.

♠ 4.7*(ii)*: This is an immediate consequence of the definition of the parallel sum, owing to Proposition 4.4*(i)* and Proposition 4.7*(i)*. □

Proof of Proposition 4.8 *(sum with identity)*.

♠ 4.8*(i)*: First, consider the assertion where the semimonotonicity holds only at a single point. Operator A is (μ_A, ρ_A)-semimonotone at $(x', y') \in \operatorname{gph} A$ if and only if $\langle x - x', y - y' \rangle \geq \mu_A \|x - x'\|^2 + \rho_A \|y - y'\|^2$ for all $(x, y) \in \operatorname{gph} A$. Defining $p := y + \alpha x$ and $p' := y' + \alpha x'$, so that $(x, p), (x', p') \in \operatorname{gph} T$, this is equivalent to

$$(1 + 2\rho_A \alpha)\langle x - x', p - p' \rangle \geq (\mu_A + \alpha(1 + \rho_A \alpha))\|x - x'\|^2 + \rho_A \|p - p'\|^2, \quad \forall (x, p) \in \operatorname{gph} T.$$

Dividing both sides by $1 + 2\rho_A \alpha > 0$, the claimed equivalence between (μ_A, ρ_A)-semimonotonicity of A at (x', y') and $\left(\frac{\mu_A + \alpha(1 + \rho_A \alpha)}{1 + 2\rho_A \alpha}, \frac{\rho_A}{1 + 2\rho_A \alpha}\right)$-semimonotonicity of T at $(x', y' + \alpha x')$ is established. The assertion for (global) semimonotonicity follows analogously by considering all points $(x', y') \in \operatorname{gph} A$.

Finally, the assertion for maximality holds since $\operatorname{gph} A$ can be enlarged with a point $(\hat{x}, \hat{y})$ without breaking (μ_A, ρ_A)-semimonotonicity if and only if $\operatorname{gph} T$ can be enlarged with $(\hat{x}, \hat{y} + \alpha\hat{x})$ without violating $\left(\frac{\mu_A + \alpha(1 + \rho_A \alpha)}{1 + 2\rho_A \alpha}, \frac{\rho_A}{1 + 2\rho_A \alpha}\right)$-semimonotonicity.

♠ 4.8*(ii)*: This result immediately follows from Proposition 4.7*(i)* and Corollary 4.6. □

Proof of Proposition 4.11 *(outer semicontinuity of a semimonotone operator)*. First, assume that $\rho\mu < \frac{1}{4}$ and let ξ and ν be defined as in (4.2). Let $(x^k, y^k) \in \operatorname{gph} A$, where $x^k \to \bar{x}$ and $y^k \to \bar{y}$. Then, it holds by construction that $x^k - \xi(y^k - \nu x^k) \in \left[(A^{-1} - \nu\mathrm{id})^{-1} - \xi\mathrm{id}\right](y^k - \nu x^k)$. Owing to Corollary 4.10, $(A^{-1} - \nu\mathrm{id})^{-1} - \xi\mathrm{id}$ is maximally monotone and thus also outer semicontinuous [77, Ex. 12.8]. Consequently, it follows that $\bar{x} - \xi(\bar{y} - \nu\bar{x}) \in \left[(A^{-1} - \nu\mathrm{id})^{-1} - \xi\mathrm{id}\right](\bar{y} - \nu\bar{x})$, which in turn implies that $\bar{x} \in A\bar{y}$, showing outer semicontinuity of A. Finally, in the case where $[\mu]_+[\rho]_+ = 1/4$, maximal monotonicity of A follows by combining Propositions 4.3*(iii)*-2. and 4.4*(ii)* and Corollary 4.6, which implies its outer semicontinuity [77, Ex. 12.8]. □

Proof of Proposition 4.12 *(resolvent of a semimonotone operator)*.

♠ 4.12*(i)*: First, assume that $\mu\rho < \frac{1}{4}$. Define $\theta := \sqrt{1 - 4\rho\mu}$ and let $\nu := \frac{2\mu}{1+\theta}$, $\xi := \frac{\rho}{\theta}$ and $\omega := \frac{\gamma}{1+\gamma\nu}$. Using the upper bound $\gamma < \frac{1+\theta}{2[-\mu]_+}$ from (4.3), it follows that $1 + \gamma\nu = 1 + \gamma\frac{2\mu}{1+\theta} \geq 1 - \frac{2\gamma[-\mu]_+}{1+\theta} > 0$. Consequently, $\omega + \xi > 0$ since

$$\frac{\gamma}{1+\gamma\nu} + \frac{\rho}{\theta} > 0 \overset{1+\gamma\nu>0}{\iff} \gamma + \left(1 + \gamma\frac{2\mu}{1+\theta}\right)\frac{\rho}{\theta} > 0 \overset{\theta>0}{\iff} \left(2\theta + \frac{4\mu\rho}{1+\theta}\right)\gamma + 2\rho > 0 \iff (1+\theta)\gamma + 2\rho > 0$$

which is satisfied by the lower bound $\gamma > \frac{2[-\rho]_+}{1+\theta}$ from (4.3). Denote $\hat{A} := (A - \nu\mathrm{id})^{-1} + \omega\mathrm{id}$. Owing to $\omega + \xi > 0$ and Minty's theorem [9, Thm. 21.1], the operator $\frac{1}{\omega+\xi}\hat{A} = \frac{1}{\omega+\xi}\left((A - \nu\mathrm{id})^{-1} - \xi\mathrm{id}\right) + \mathrm{id}$ has full range iff $\frac{1}{\omega+\xi}((A - \nu\mathrm{id})^{-1} - \xi\mathrm{id})$ is maximally monotone. Since $\omega + \xi > 0$, this implies that $\hat{A}$ has full range iff the operator $\tilde{A} := (A - \nu\mathrm{id})^{-1} - \xi\mathrm{id}$ is maximally monotone.

On the other hand, by definition $1 + 2\xi\nu = 1 + \frac{1-\theta^2}{\theta(1+\theta)} = \frac{1}{\theta} > 0$. Hence, using Corollary 4.9*(i)*, the operator $\tilde{A} := (A - \nu\mathrm{id})^{-1} - \xi\mathrm{id}$ is maximally monotone if and only if A is maximally $\left(\frac{\nu(1+\xi\nu)}{1+2\xi\nu}, \frac{\xi}{1+2\xi\nu}\right) = (\mu, \rho)$ semimonotone, where the equality is the result of substituting ξ and ν. Therefore, it suffices to show $\hat{A}$ has full range iff $\mathcal{R}(\mathrm{id} + \gamma A) = \mathbb{R}^n$, i.e., that the resolvent has full domain. First suppose that $\hat{A}$ has full range. Fix $y \in \mathbb{R}^n$. Then, by the full range assumption for $\hat{A}$, there exists some $x \in \mathbb{R}^n$ such that $\frac{\omega}{\gamma}y \in (A - \nu\mathrm{id})^{-1}x + \omega x$. Letting $z = \omega(\frac{1}{\gamma}y - x)$ we have

$$\frac{1}{\gamma}y - \frac{1}{\omega}z \in (A - \nu\mathrm{id})z \iff \frac{1}{\gamma}y \in (A + (\frac{1}{\omega} - \nu)\mathrm{id})z \iff y \in (\gamma A + \mathrm{id})z, \quad (\text{B.7})$$

where in the last equivalence we used the fact that $\gamma = \frac{\omega}{1-\omega\nu}$. Since the choice of y was arbitrary, $J_{\gamma A}$ has full domain. The converse follows similarly. Assume that $\mathcal{R}(I + \gamma A) = \mathbb{R}^n$. For any fixed $y \in \mathbb{R}^n$ there exists some $z \in \mathbb{R}^n$ such that $y \in \gamma A z + z$. Therefore, by (B.7) and letting $x = \frac{1}{\gamma}y - \frac{1}{\omega}z$ we obtain $\frac{\omega}{\gamma}y \in \hat{A}x$, thus establishing that $\mathcal{R}(\hat{A}) = \mathbb{R}^n$.

Finally, consider the case $[\mu]_+[\rho]_+ = \frac{1}{4}$. The range for the stepsize in (4.3) then reduces to $\gamma \in (0, +\infty)$. First, suppose that $J_{\gamma A}$ has full domain, i.e., $\mathcal{R}(\mathrm{id} + \gamma A) = \mathbb{R}^n$. Fix $v \in \mathbb{R}^n$. Then, there exists $(x, y) \in \mathrm{gph}\, A$ such that $v = x + \gamma y$. By Proposition 4.3*(iii)*-1., $y = 2\mu x + c$ for some $c \in \mathbb{R}$ and some $x \in \mathrm{dom}\, A$, and therefore, $v = (1 + 2\gamma\mu)x + c\gamma$. Since $1 + 2\gamma\mu > 0$ and the choice of v was arbitrary, necessarily $\mathrm{dom}\, A = \mathbb{R}^n$, and as such A is maximally (μ, ρ)-semimonotone, see Proposition 4.3*(iii)*-2.. Now, suppose that A is maximally (μ, ρ)-semimonotone. Then, by Proposition 4.3*(iii)*-2. it holds that $A : x \mapsto 2\mu x + c$ for some $c \in \mathbb{R}^n$, which implies that A is maximally monotone. Therefore by Minty's theorem [9, Thm. 21.1] the resolvent $J_{\gamma A}$ has full domain.

♠ 4.12*(ii)*: Take $(u, v), (u', v') \in \mathrm{gph}\, J_{\gamma A}$. By semimonotonicity of A and using the fact that $\frac{1}{\gamma}(u - v) \in Av$, $\frac{1}{\gamma}(u' - v') \in Av'$, we have

$$\mu\|v - v'\|^2 + \rho\gamma^{-2}\|(u - v) - (u' - v')\|^2 \le \gamma^{-1}\langle(u - v) - (u' - v'), v - v'\rangle.$$

Expanding the last term of the left-hand side and rearranging, we get

$$\left(1 + \mu\gamma + \rho\gamma^{-1}\right)\|v - v'\|^2 + \rho\gamma^{-1}\|u - u'\|^2 \le \left(1 + 2\rho\gamma^{-1}\right)\langle u - u', v - v'\rangle$$
$$\le \left|1 + 2\rho\gamma^{-1}\right|\|u - u'\|\|v - v'\|. \tag{B.8}$$

By Fact A.3*(ii)* it follows that (4.3) provides all strictly positive solutions for γ of $1 + \mu\gamma + \rho\gamma^{-1} > 0$. Hence, if $u = u'$ then necessarily $v = v'$, and the Lipschitz inequality holds trivially. Else, divide both sides by $\|u - u'\|^2$ to obtain

$$\left(1 + \mu\gamma + \rho\gamma^{-1}\right)\omega^2 - \left|1 + 2\rho\gamma^{-1}\right|\omega + \rho\gamma^{-1} \le 0,$$

where $\omega := \frac{\|v-v'\|}{\|u-u'\|}$. Solving this inequality for ω yields the claimed Lipschitz modulus. $\qquad\square$

B.3 Examples of functions with semimonotone subdifferentials

Proof of Proposition 4.14 *(functions with upper and lower curvature).*

♠ 4.14*(i)*: Follows by definition.

♠ 4.14*(ii)*: By [77, Prop. 12.60] and strong convexity of $-f$, the conjugate $(-f)^*$ is $-\frac{1}{\ell}$-smooth, i.e.

$$\langle\nabla(-f)^*(x) - \nabla(-f)^*(x'), x - x'\rangle \le -\tfrac{1}{\ell}\|x - x'\|^2, \quad \forall x, x' \in \mathbb{R}^n. \tag{B.9}$$

Moreover, since $-f$ is continuous and convex, we have by [77, Prop. 11.3] that $\nabla(-f)^* = (\partial(-f))^{-1}$, so that (B.9) is equivalent to

$$\langle y - y', x - x'\rangle \le -\tfrac{1}{\ell}\|y - y'\|^2, \quad \forall(x, y), (x', y') \in \mathrm{gph}\, \partial(-f). \tag{B.10}$$

Finally, since $-f$ is finite and convex, it follows from [77, Cor. 9.21] that $\partial f(\bar{x}) \subseteq -\partial(-f)(\bar{x})$, $\forall\bar{x} \in \mathbb{R}^n$, so that

$$\langle y - y', x - x'\rangle \ge \tfrac{1}{\ell}\|y - y'\|^2, \quad \forall(x, y), (x', y') \in \mathrm{gph}\, \partial f \subseteq -\mathrm{gph}\, \partial(-f).$$

♠ 4.14*(iii)*: Consider the function $\varphi = f - \frac{\alpha}{2}\|\cdot\|^2$. Then, by definition $\varphi \in \mathscr{F}^{\ell-\alpha}(\mathbb{R}^n)$, implying that $\partial\varphi$ is $(0, \frac{1}{\ell-\alpha})$-semimonotone owing to Proposition 4.14*(ii)* since $\alpha > \ell$. Since $\partial f = \partial\varphi + \alpha\mathrm{id}$, [72, Ex. 8.8(c)], it follows from Proposition 4.8 that for $c > \frac{1}{\alpha-\ell}$ the claimed semimonotonicity holds.

♠ 4.14*(iv)*, 4.14*(v)*: Suppose that $\ell > \sigma$, since if $\ell = \sigma$, then, $\nabla f = \ell\mathrm{id}$ and the claims follow trivially. Consider the function $\varphi = f - \frac{\sigma}{2}\|\cdot\|^2 \in \mathscr{F}_0^{\ell-\sigma}(\mathbb{R}^n)$. Since φ is finite and convex its subdifferential $\partial\varphi(x)$ is nonempty for all $x \in \mathbb{R}^n$ [72, Thm. 23.4]. Consequently, φ is continuously differentiable due to [84, Lem. 2.1] and $\nabla\varphi$ is $(0, \frac{1}{\ell-\sigma})$-semimonotone due to [60, Thm. 2.1.10]. Therefore, by Proposition 4.4*(i)*, there exists a monotone mapping $\tilde{A}$ for which $(\nabla\varphi)^{-1} = \tilde{A} + \frac{1}{\ell-\sigma}$ and thus, owing to $\nabla f = \nabla\varphi + \sigma\mathrm{id}$ [77, Ex. 8.8(c)] it holds that

$$\nabla f = \nabla\varphi + \sigma\mathrm{id} = (\tilde{A} + \tfrac{1}{\ell-\sigma}\mathrm{id})^{-1} + \sigma\mathrm{id}.$$

The claimed results from Propositions 4.14*(iv)* and 4.14*(v)* then follow from Corollary 4.9, where we let $\xi = \frac{1}{\ell-\sigma}$ and $\nu = \sigma$. This is because $\ell - \sigma > 0$ and therefore $1 + 2\xi\nu > 0$ (resp. ≤ 0) if and only if $\sigma + \ell > 0$ (resp. ≤ 0). $\qquad\square$

Proof of Corollary 4.15 (*pointwise minimum*). Since $f_i \in \mathscr{F}^{\ell_{f_i}}(\mathbb{R}^n)$, it follows that $\frac{\bar{\ell}}{2}\|\cdot\|^2 - f_i$ is convex. Therefore, $\frac{\bar{\ell}}{2}\|x\|^2 - f(x) = \max\left\{\frac{\bar{\ell}}{2}\|x\|^2 - f_1(x), \dots, \frac{\bar{\ell}}{2}\|x\|^2 - f_N(x)\right\}$ is convex being the pointwise max of convex functions and thus $f \in \mathscr{F}^{\bar{\ell}}(\mathbb{R}^n)$. The claimed result then follows immediately from Propositions 4.14*(ii)* and 4.14*(iii)*. $\square$

Proof of Proposition 4.16 (*infimal convolution*).

♠ 4.16*(i)*: Let $f(w) = f_1(w_1) + f_2(w_2)$, and L be the linear mapping $Lw = w_1 + w_2$. Then, the infimal convolution may be viewed as the parametric minimization of $\tilde{f}(w, s) = f(w) + \delta_{\{0\}}(Lw - s)$ with respect to the variable w, i.e., $\varphi(s) = \min_w \tilde{f}(w, s)$. Let $M(s) := \arg\min_{(w_1, w_2)} \{f_1(w_1) + f_2(w_2) \mid w_1 + w_2 = s\}$. Then, owing to (4.4) it holds by [77, Thm. 10.13] for all $\bar{s} \in \mathbb{R}^n$ that

$$\partial\varphi(\bar{s}) \subseteq \bigcup_{\bar{x} \in M(\bar{s})} \left\{y \in R^n \mid (0, y) \in \partial\tilde{f}(\bar{x}, \bar{s})\right\}. \tag{B.11}$$

Consider $\bar{f}(w, u) = f(w) + \delta_{\{0\}}(u)$ and the linear transformation $K : (w, s) \mapsto (w, Lw - s)$. Then, $\tilde{f} = \bar{f} \circ K$. Since K has full rank, by [77, Ex. 10.7] for any pair $(\bar{x}, \bar{s})$ such that $\bar{x} \in M(\bar{s})$ it holds that

$$\partial\tilde{f}(\bar{x}, \bar{s}) = (K^\top \circ \partial\bar{f} \circ K)(\bar{x}, \bar{s}) \quad \text{where} \quad \partial\bar{f}(w, u) = (\partial f(w), N_{\{0\}}(u)).$$

Using this in (B.11) yields

$$\partial\varphi(\bar{s}) \subseteq \bigcup_{\bar{x} \in M(\bar{s})} \left\{y \in R^n \mid (0, y) \in \left(\partial f(\bar{x}) + L^\top N_{\{0\}}(L\bar{x} - \bar{s}), -N_{\{0\}}(L\bar{x} - \bar{s})\right)\right\} = \bigcup_{(\bar{x}_1, \bar{x}_2) \in M(\bar{s})} \partial f_1(\bar{x}_1) \cap \partial f_2(\bar{x}_2).$$

Hence, the claimed result follows from [9, Prop. 25.30(i)].

♠ 4.16*(ii)*: Follows immediately from Proposition 4.16*(i)* and Proposition 4.7*(ii)*. $\square$

Proof of Proposition 4.17 (*saddle operator*). Consider the saddle envelope

$$\varphi_\delta(x, y) := \min_{u \in \mathbb{R}^n} \max_{v \in \mathbb{R}^m} \{\varphi(u, v) + \tfrac{1}{2\delta}\|u - x\|^2 - \tfrac{1}{2\delta}\|v - y\|^2\},$$

along with the associated resolvent $J_{\delta T_\varphi} := (\mathrm{id} + \delta T_\varphi)^{-1}$. For any $(z, w) \in \mathrm{gph}\, J_{\delta T_\varphi}$ with the decomposition $z = (x, y)$ and $w = (u, v)$ we have by construction that

$$0 = \nabla_x\varphi(w) + \tfrac{1}{\delta}(u - x) = \nabla_x\varphi(w) - \nabla_x\varphi_\delta(z) \quad \text{and} \quad 0 = \nabla_y\varphi(w) - \tfrac{1}{\delta}(v - y) = \nabla_y\varphi(w) - \nabla_y\varphi_\delta(z).$$

Hence,

$$T_\varphi(w) = (\nabla_x\varphi(w), -\nabla_y\varphi(w)) = (\nabla_x\varphi_\delta(z), -\nabla_y\varphi_\delta(z)) =: T_{\varphi_\delta}(z). \tag{B.12}$$

On the other hand, it follows from [40, Prop. 1] that

$$\tfrac{1}{\delta}\mathrm{I} \succeq \nabla_{xx}\varphi_\delta(z) \succeq \tfrac{\alpha}{1+\delta\alpha}\mathrm{I}, \quad \tfrac{1}{\delta}\mathrm{I} \succeq -\nabla_{yy}\varphi_\delta(z) \succeq \tfrac{\alpha}{1+\delta\alpha}\mathrm{I}, \quad \forall z, \tag{B.13}$$

thus implying that the operator T_{φ_δ} is $\frac{\alpha}{1+\delta\alpha}$-monotone. Let $w_1, w_2 \in \mathbb{R}^n$ and $z_i = w_i + \delta T_\varphi w_i$ for $i = 1, 2$. Then, $(z_i, w_i) \in \mathrm{gph}\, J_{\delta T_\varphi}$ for $i = 1, 2$ and we have that

$$\langle T_\varphi w_1 - T_\varphi w_2, w_1 - w_2 \rangle = \langle T_\varphi w_1 - T_\varphi w_2, z_1 - z_2 \rangle - \delta\|T_\varphi w_1 - T_\varphi w_2\|^2$$
$$\overset{(B.12)}{=} \langle T_{\varphi_\delta} z_1 - T_{\varphi_\delta} z_2, z_1 - z_2 \rangle - \delta\|T_\varphi w_1 - T_\varphi w_2\|^2$$
$$\overset{(B.13)}{\geq} \tfrac{\alpha}{1+\alpha\delta}\|z_1 - z_2\|^2 - \delta\|T_\varphi w_1 - T_\varphi w_2\|^2. \tag{B.14}$$

The second assertion follows from the above inequality noting that $1 + \alpha\delta > 0$. Moreover, plugging $z_i = w_i + \delta T_\varphi w_i$ in (B.14) and multiplying by $(1 + \alpha\delta)$ yields

$$(1 - \alpha\delta)\langle T_\varphi w_1 - T_\varphi w_2, w_1 - w_2 \rangle \geq \alpha\|w_1 - w_2\|^2 - \delta\|T_\varphi w_1 - T_\varphi w_2\|^2,$$

establishing the first assertion. $\square$

B.4 Douglas-Rachford splitting for semimonotone operators

Proof of Lemma 5.1 *(primal-dual operator).* Let $(x_A, y_A) \in \mathrm{gph}\, A$ and $(x_B, y_B) \in \mathrm{gph}\, B$. Additionally, if Assumption IV holds, let $(x'_A, y'_A) \in \mathrm{gph}\, A$ and $(x'_B, y'_B) \in \mathrm{gph}\, B$. Otherwise, if Assumption III holds, let $(x^\star, y^\star) \in S^\star$ and define $(x'_A, y'_A) := (x^\star, -y^\star) \in \mathrm{gph}\, A$ and $(x'_B, y'_B) := (x^\star, y^\star) \in \mathrm{gph}\, B$. Then, by definition of the primal-dual operator from (PD-I), it holds that

$$(z, v) := \big((x_A, y_B), (y_A + y_B, x_B - x_A)\big) \in \mathrm{gph}\, T_{\mathrm{PD}} \quad \text{and} \quad (z', v') := \big((x'_A, y'_B), (y'_A + y'_B, x'_B - x'_A)\big) \in \mathrm{gph}\, T_{\mathrm{PD}}.$$

Owing to the semimonotonicity assumptions and Proposition 4.4(i), this implies that

$$\begin{aligned}
\langle v - v', z - z' \rangle &= \langle y_A + y_B - (y'_A + y'_B), x_A - x'_A \rangle + \langle x_B - x_A - (x'_B - x'_A), y_B - y'_B \rangle \\
&= \langle y_A - y'_A, x_A - x'_A \rangle + \langle x_B - x'_B, y_B - y'_B \rangle \\
&\geq \mu_A \|x_A - x'_A\|^2 + \rho_A \|y_A - y'_A\|^2 + \rho_B \|y_B - y'_B\|^2 + \mu_B \|x_B - x'_B\|^2.
\end{aligned} \tag{B.15}$$

On the other hand, it follows from the Cauchy-Schwarz inequality that

$$\begin{aligned}
q_V(v - v') &= \beta_{\mathrm{P}} \|y_A + y_B - (y'_A + y'_B)\|^2 + \beta_{\mathrm{D}} \|x_B - x_A - (x'_B - x'_A)\|^2 \\
&\leq \beta_{\mathrm{P}} \|y_A - y'_A\|^2 + \beta_{\mathrm{P}} \|y_B - y'_B\|^2 + 2|\beta_{\mathrm{P}}| \|y_A - y'_A\| \|y_B - y'_B\| \\
&\quad + \beta_{\mathrm{D}} \|x_B - x'_B\|^2 + \beta_{\mathrm{D}} \|x_A - x'_A\|^2 + 2|\beta_{\mathrm{D}}| \|x_B - x'_B\| \|x_A - x'_A\|.
\end{aligned} \tag{B.16}$$

Combining (B.15) and (B.16), it follows that the primal-dual operator T_{PD} has V-oblique weak Minty solutions at $S^\star$ if Assumption III holds and that it is V-comonotone if Assumption IV holds, provided that $X := \left(\begin{smallmatrix} \mu_A - \beta_{\mathrm{D}} & -|\beta_{\mathrm{D}}| \\ -|\beta_{\mathrm{D}}| & \mu_B - \beta_{\mathrm{D}} \end{smallmatrix}\right) \succeq 0$ and $Y := \left(\begin{smallmatrix} \rho_B - \beta_{\mathrm{P}} & -|\beta_{\mathrm{P}}| \\ -|\beta_{\mathrm{P}}| & \rho_A - \beta_{\mathrm{P}} \end{smallmatrix}\right) \succeq 0$. As previously discussed in the proof of Proposition 4.7(i), the largest possible value for β_{D} satisfying $X \succeq 0$ is given by $\mu_A \,\square\, \mu_B$ if either $\mu_A + \mu_B > 0$ or $\mu_A = \mu_B = 0$ (and analogously the largest value for β_{P} satisfying $Y \succeq 0$ is given by $\rho_A \,\square\, \rho_B$ if either $\rho_A + \rho_B > 0$ or $\rho_A = \rho_B = 0$), which completes the proof for Lemma 5.1(i). To show Lemma 5.1(ii), it only remains to show maximality. To this end, consider an arbitrary stepsize γ complying with Table 3, where γ_- and γ_+ are defined as in (5.2). Then, by Lemma A.6 the resolvents $J_{\gamma A}$ and $J_{\gamma B}$ have full domain, which in turn implies that the preconditioned resolvent $(P + T_{\mathrm{PD}})^{-1} P$ has full domain, where P is given by (3.1) [77, Lem. 12.14]. The claim then follows from Proposition A.2. □

Proof of Theorem 5.2 *(convergence of DRS under semimonotonicity).* Consider the following implications.

♠ For Theorem 5.2(i), it follows from Assumption III and Lemma 5.1(i) that Assumption II.A3 holds.

♠ For Theorem 5.2(ii), it follows from Assumption IV and Lemma A.6 that Assumptions II.A1 and II.A2 hold and from Assumption IV and Lemma 5.1(ii) that T_{PD} is maximally $V := \mathrm{blkdiag}(\beta_{\mathrm{P}} I_n, \beta_{\mathrm{D}} I_n)$-comonotone, where $[\beta_{\mathrm{D}}]_- [\beta_{\mathrm{P}}]_- < \frac{1}{4}$.

Moreover, the stepsize γ satisfies (3.3), since the provided bounds on γ from Table 3 are obtained by plugging in (5.1) into (3.3). Based on these observations, the claims follow directly from Theorem 3.3 and Corollary 3.4, concluding the proof. □

C Examples

Details of Example 2.6 *(toy example).*

♠ 2.6(i)-1.: Since T is equal to the linear mapping $T = \left[\begin{smallmatrix} b & a \\ -a & b \end{smallmatrix}\right]$, it is ρ-comonotone if and only if $\frac{1}{2}(T + T^\top) \preceq \rho T^\top T$ which holds for $\rho \leq b/a^2 + b^2$.

♠ 2.6(i)-2.: To show that the result from Theorem 2.4(v) is tight, define

$$H := I_2 + \lambda\big((I_2 + \gamma T)^{-1} - I_2\big) = I_2 + \lambda\left(\begin{bmatrix} 1 + \gamma b & \gamma a \\ -\gamma a & 1 + \gamma b \end{bmatrix}^{-1} - I_2\right) = I_2 + \lambda\left(\tfrac{1}{(1+\gamma b)^2 + \gamma^2 a^2}\begin{bmatrix} 1 + \gamma b & -\gamma a \\ \gamma a & 1 + \gamma b \end{bmatrix} - I_2\right),$$

such that the iteration of the proximal point algorithm with fixed relaxation parameter λ can be expressed as the linear dynamical system $z^{k+1} = Hz^k$. This system is globally asymptotically stable if and only if the spectral radius of H, given by $\sqrt{1 - \tfrac{\lambda(2\bar{\lambda} - \lambda)(a^2 + b^2)\gamma^2}{(a^2 + b^2)\gamma^2 + 2b\gamma + 1}}$, is strictly less than one, which holds iff (2.18) is satisfied.

♠ 2.6(ii)-1.: Recall that for a given $\rho \in \mathbb{R}$, the ρ-weak Minty solutions of this problem are given by the set $\mathcal{S}^\star \subseteq \operatorname{zer} T$ for which it holds that

$$\langle T(z), z - z^\star \rangle \geq \rho\|T(z)\|^2, \quad \forall z \in \mathbb{R}^2, z^\star \in \mathcal{S}^\star. \tag{C.1}$$

The analytical expression for f as defined in Figure 3 is given by

$$f\big(\|z\|\big) := \begin{cases} \|z\|, & \text{if } \|z\| \leq 2/5, \\ 4/5 - \|z\|, & \text{if } \|z\| \in (2/5, 4/5), \\ 0, & \text{if } \|z\| \in [4/5, 1], \\ 5/2(\|z\| - 1), & \text{if } \|z\| \in (1, 7/5), \\ 1, & \text{if } \|z\| \geq 7/5. \end{cases} \tag{C.2}$$

Therefore, by definition of T, it holds that $\operatorname{zer} T = \big\{z \in \mathbb{R}^2 \mid \|z\| \in \{0\} \cup [4/5, 1]\big\}$ and that (C.1) corresponds to

$$f\big(\|z\|\big)\big(b\|z\|^2 - b(z_1 z_1^\star + z_2 z_2^\star) + a(z_1 z_2^\star - z_1^\star z_2)\big) \geq \rho(a^2 + b^2)f\big(\|z\|\big)^2\|z\|^2, \qquad \forall z \in \mathbb{R}^2, z^\star \in \mathcal{S}^\star. \tag{C.3}$$

This obviously holds for all $z \in \operatorname{zer} T$. For $z \notin \operatorname{zer} T$, we distinguish between the following two cases.

⋄ $\|z^\star\| = 0$: Then, (C.3) holds iff for all $z \notin \operatorname{zer} T$ it holds that $\rho \leq \tfrac{b}{(a^2+b^2)f(\|z\|)}$. Since $f(\|z\|) \in (0, 1]$, this condition is satisfied for $\rho = b/a^2+b^2$.

⋄ $\|z^\star\| \in [4/5, 1]$: Let $z = \epsilon z^\star$ where $\epsilon \in (0, 2/5)$. Then, $f(\|z\|) = \epsilon\|z^\star\|$ and (C.3) holds iff $\rho \leq \tfrac{b(1-\epsilon^{-1})}{\epsilon(a^2+b^2)\|z^\star\|}$ Taking the limit for ϵ to zero from above, we get that $\lim_{\epsilon \to 0+} \tfrac{b(1-\epsilon^{-1})}{\epsilon(a^2+b^2)\|z^\star\|} = -\infty$ and thus (C.3) cannot be satisfied for any $\rho \in \mathbb{R}$. □

Details of *Example 2.7* *(Von Neumann's economic equilibrium model).* We proceed by establishing three intermediate facts for the set $C := \Delta_2 \times \Delta_2$ and the operators F and N_C. First, observe that by definition of the simplex

$$z \in C \iff z = (s, 1 - s, t, 1 - t) \text{ for some } s, t \in [0, 1]. \tag{C.4}$$

Second, recall that

$$F(z) = (F_1(z), F_2(z), F_3(z), F_4(z)) = \big(\nabla_x f(x, y), -\nabla_y f(x, y)\big) = \Big(\tfrac{\langle x, Sy\rangle Ry - \langle x, Ry\rangle Sy}{\langle x, Sy\rangle^2}, \tfrac{\langle x, Ry\rangle S^\top x - \langle x, Sy\rangle R^\top x}{\langle x, Sy\rangle^2}\Big).$$

Therefore, using the parametrization from (C.4) it holds for any $z \in C$ that

$$\langle F(z), z - z^\star \rangle = \tfrac{\epsilon(s+t) - st(\epsilon+s)}{1/2(2-s)^2}, \quad F_1(z) - F_2(z) = \tfrac{\epsilon - t(2+\epsilon/2)}{1/2(2-s)^2} \quad \text{and} \quad F_3(z) - F_4(z) = \tfrac{\epsilon/2 + s}{1/2(2-s)}. \tag{C.5}$$

Finally, denoting the vector of all 1s of dimension n by $\mathbf{1}_n$, it holds owing to [43, Ex. 5.2.6(c)] that

$$N_C(z) = \begin{cases} \big\{\alpha \otimes \mathbf{1}_2 - \beta \mid \alpha \in \mathbb{R}^2, \beta \in \mathbb{R}_+^4, \langle \beta, z \rangle = 0\big\}, & \text{if } z \in C, \\ \emptyset, & \text{otherwise.} \end{cases} \tag{C.6}$$

We now proceed with the main argument. In [22, Prop. 2(1)], it was shown that $z^\star = (0, 1, 0, 1)$ is the unique Nash equilibrium of (2.19) (noting that $f(x, y) - 2 = \langle x, \begin{bmatrix} -1 & \epsilon/2 \\ -\epsilon/2 & 0 \end{bmatrix} y\rangle/\langle x, Sy\rangle$). This Nash equilibrium is a ρ_ϵ-weak
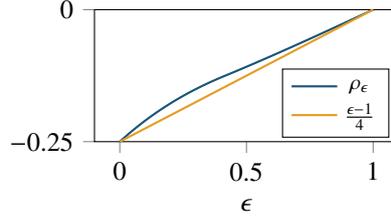
Figure 9: *Comparison between the optimal weak Minty constant ρ_ϵ and its lower bound $\frac{\epsilon-1}{4}$ for Example 2.7.*

Minty solution of T provided that $\rho_\epsilon = \inf_{z \in C} \varrho_\epsilon(z)$ where $\varrho_\epsilon(z) = \inf_{v \in N_C(z)} \langle F(z)+v, z-z^\star \rangle / \|F(z)+v\|^2$. From now onwards, we tacitly assume that $z \in C$. Then, owing to (C.6), it holds that

$$\varrho_\epsilon(z) = \inf\left\{\frac{\langle F(z), z-z^\star \rangle + \langle \alpha \otimes \mathbf{1}_2 - \beta, z-z^\star \rangle}{\|F(z)+\alpha \otimes \mathbf{1}_2-\beta\|^2} \mid \alpha \in \mathbb{R}^2, \beta \in \mathbb{R}^4_+, \langle \beta, z \rangle = 0\right\}.$$

Owing to the parametrization from (C.4) and the fact that $z^\star = (0, 1, 0, 1)$, it holds that $\langle \alpha \otimes \mathbf{1}_2, z - z^\star \rangle = 0$ and $\langle \beta, z^\star \rangle = \beta_2 + \beta_4$, so that

$$\varrho_\epsilon(z) = \inf\left\{\frac{\langle F(z), z-z^\star \rangle + \beta_2 + \beta_4}{\|F(z)+\alpha \otimes \mathbf{1}_2-\beta\|^2} \mid \alpha \in \mathbb{R}^2, \beta \in \mathbb{R}^4_+, \langle \beta, z \rangle = 0\right\}.$$

Note that by construction the sign of $\varrho_\epsilon(z)$ is equal to the sign of $\langle F(z), z - z^\star \rangle$: if $\langle F(z), z - z^\star \rangle$ is positive, this is obvious since $\beta_i \geq 0$; if it is negative, then this can be seen by taking the candidate solution $\beta_2 = \beta_4 = 0$. Therefore, it follows from (C.5) that $\varrho_\epsilon(z) < 0$ if and only if $\epsilon(s + t) - st(\epsilon + s) < 0$, i.e., if and only if z lies in the nonempty set

$$D := \left\{(s, 1 - s, t, 1 - t) \mid s \in (\sqrt{\epsilon}, 1], \ t \in (\bar{t}(s), 1]\right\} \quad \text{where} \quad \bar{t}(s) := \frac{\epsilon s}{s(\epsilon+s)-\epsilon}.$$

Therefore, $\rho_\epsilon = \inf_{z \in C} \varrho_\epsilon(z) = \inf_{z \in D} \varrho_\epsilon(z)$. Consider $z \in D \subset C$. Then, by construction $\varrho_\epsilon(z)$ is negative, which implies that $\varrho_\epsilon(z)$ is minimized for α that minimizes the denominator, given by

$$\underset{\alpha \in \mathbb{R}^2}{\arg\min} \|F(z) + \alpha \otimes \mathbf{1}_2 - \beta\|^2 = \left(\tfrac{1}{2}(F_1(z) - F_2(z) - \beta_1 + \beta_2), \tfrac{1}{2}(F_3(z) - F_4(z) - \beta_3 + \beta_4)\right),$$

this implies that

$$\varrho_\epsilon(z) = \inf\left\{\frac{2\langle F(z), z-z^\star \rangle + 2\beta_2 + 2\beta_4}{\left(F_1(z)-\beta_1-F_2(z)+\beta_2\right)^2 + \left(F_3(z)-\beta_3-F_4(z)+\beta_4\right)^2} \mid \beta \in \mathbb{R}^4_+, \langle \beta, z \rangle = 0\right\}.$$

Plugging in the parametrization for z from (C.4), using (C.5) and introducing $\bar{\beta} := 1/2(2 - s)^2\beta$, it follows that

$$\varrho_\epsilon(s, t) = \inf\left\{\frac{4(2-s)^2\left(\epsilon(s+t)-st(\epsilon+s)+\bar{\beta}_2+\bar{\beta}_4\right)}{\left(2\epsilon-t(4+\epsilon)-2\bar{\beta}_1+2\bar{\beta}_2\right)^2+\left((\epsilon+2s)(2-s)-2\bar{\beta}_3+2\bar{\beta}_4\right)^2} \mid \bar{\beta} \in \mathbb{R}^4_+, \langle \bar{\beta}, (s, 1-s, t, 1-t) \rangle = 0\right\},$$

We claim that $\bar{\beta} = 0$ at the infimum for any $s \in (\sqrt{\epsilon}, 1]$ and $t \in (\bar{t}(s), 1]$. For $\bar{\beta}_1$ and $\bar{\beta}_3$, this follows immediately from the equality constraint and the fact that s and t are strictly positive, so that

$$\varrho_\epsilon(s, t) = \inf\left\{\frac{4(2-s)^2\left(\epsilon(s+t)-st(\epsilon+s)+\bar{\beta}_2+\bar{\beta}_4\right)}{\left(2\epsilon-t(4+\epsilon)+2\bar{\beta}_2\right)^2+\left((\epsilon+2s)(2-s)+2\bar{\beta}_4\right)^2} \mid (\bar{\beta}_2, \bar{\beta}_4) \in \mathbb{R}^2_+, \langle (\bar{\beta}_2, \bar{\beta}_4), (1-s, 1-t) \rangle = 0\right\}.$$

If $s \in (\sqrt{\epsilon}, 1)$ and $t \in (\bar{t}(s), 1)$, the claim for $\bar{\beta}_2$ and $\bar{\beta}_4$ also follows from the equality constraint. Otherwise, consider the following cases.

♠ If $s = 1$ and $t \in (\bar{t}(1), 1) = (\epsilon, 1)$, then $\bar{\beta}_4 = 0$ due to the equality constraint and

$$\varrho_\epsilon(1, t) = \inf_{\bar{\beta}_2 \in \mathbb{R}_+} \frac{4(\epsilon-t+\bar{\beta}_2)}{(2\epsilon-t(4+\epsilon)+2\bar{\beta}_2)^2+(\epsilon+2)^2} = \inf_{\bar{\beta}_2 \in [0,t-\epsilon)} \frac{4(\epsilon-t+\bar{\beta}_2)}{(2\epsilon-t(4+\epsilon)+2\bar{\beta}_2)^2+(\epsilon+2)^2}.$$

Since this function is monotonically increasing for $\bar{\beta}_2 \in [0, t - \epsilon)$, the infimum is attained for $\bar{\beta}_2 = 0$.

37

♠ If $s \in (\sqrt{\epsilon}, 1)$ and $t = 1$, then $\bar{\beta}_2 = 0$ due to the equality constraint and

$$\varrho_\epsilon(s, 1) = \inf_{\bar{\beta}_4 \in \mathbb{R}_+} \frac{4(2-s)^2(\epsilon-s^2+\bar{\beta}_4)}{(\epsilon-4)^2+\left((\epsilon+2s)(2-s)+2\bar{\beta}_4\right)^2} = \inf_{\bar{\beta}_4 \in [0, s^2-\epsilon)} \frac{4(2-s)^2(\epsilon-s^2+\bar{\beta}_4)}{(\epsilon-4)^2+\left((\epsilon+2s)(2-s)+2\bar{\beta}_4\right)^2}.$$

Since this function is monotonically increasing for $\bar{\beta}_4 \in [0, s^2 - \epsilon)$, the infimum is attained for $\bar{\beta}_4 = 0$.

♠ If $s = 1$ and $t = 1$, then

$$\varrho_\epsilon(1, 1) = \inf_{(\bar{\beta}_2, \bar{\beta}_4) \in \mathbb{R}_+^2} \frac{4(\epsilon-1+\bar{\beta}_2+\bar{\beta}_4)}{(\epsilon-4+2\bar{\beta}_2)^2+(\epsilon+2+2\bar{\beta}_4)^2},$$

of which the global minimum is attained for $\bar{\beta}_2 = \bar{\beta}_4 = 0$.

Having established that $\bar{\beta} = 0$ at the infimum, it follows that

$$\rho_\epsilon = \inf\left\{\varrho_\epsilon(s, t) \mid s \in (\sqrt{\epsilon}, 1], \ t \in \left(\frac{\epsilon s}{s(\epsilon+s)-\epsilon}, 1\right]\right\} \quad \text{where} \quad \varrho_\epsilon(s, t) = \frac{4(2-s)^2\left(\epsilon(s+t)-st(\epsilon+s)\right)}{\left(2\epsilon-t(4+\epsilon)\right)^2+(\epsilon+2s)^2(2-s)^2}. \tag{C.7}$$

Finally, the lower bound $\frac{\epsilon-1}{4} \leq \rho_\epsilon$ is obtained by solving this parametric optimization problem numerically (see also Figure 9). $\qquad\square$

Details of Example 6.1 (saddle point problem).

♠ 6.1*(i)*: By defining $\bar{H} := (\mathrm{I} + \gamma A)^{-1}$, $\widehat{H} := (\mathrm{I} + \gamma B)^{-1}(2\bar{H} - \mathrm{I})$ and $H := \mathrm{I} + \lambda(\widehat{H} - \bar{H})$, DRS can be expressed as the linear dynamical system $s^{k+1} = Hs^k$. This system is globally asymptotically stable if and only if the spectral radius of H is strictly less than one, which through elementary algebra can be shown to hold iff (6.2) is satisfied.

♠ 6.1*(ii)*: The primal-dual operator and its inverse are given by

$$T_{\mathrm{PD}} = \begin{bmatrix} A & \mathrm{I}_2 \\ -\mathrm{I}_2 & B^{-1} \end{bmatrix} \quad \text{and} \quad T_{\mathrm{PD}}^{-1} = \begin{bmatrix} (A+B)^{-1} & -(A+B)^{-1}B \\ B(A+B)^{-1} & B-B(A+B)^{-1}B \end{bmatrix}$$

where we used the Schur complement lemma. Consequently, when in (2.1) the vector v is restricted to $\mathcal{R}(P)$ we may equivalently state Assumption II.A3 as

$$z^\top\left(\frac{T_{\mathrm{PD}}+T_{\mathrm{PD}}^\top}{2} - T_{\mathrm{PD}}^\top V T_{\mathrm{PD}}\right)z \geq 0, \qquad \text{for all } z \in \mathbb{R}^n : z \in T_{\mathrm{PD}}^{-1}\mathcal{R}(P), \tag{C.8}$$

where V is defined as in (3.2). Using that $U = \begin{bmatrix} \mathrm{I}_2 \\ -\gamma\mathrm{I}_2 \end{bmatrix}$ is an orthogonal basis for $\mathcal{R}(P)$, it follows that (C.8) is satisfied iff

$$(T_{\mathrm{PD}}^{-1}U)^\top\left(\frac{T_{\mathrm{PD}}+T_{\mathrm{PD}}^\top}{2} - T_{\mathrm{PD}}^\top V T_{\mathrm{PD}}\right)T_{\mathrm{PD}}^{-1}U \geq 0 \quad \Longleftrightarrow \quad \frac{1}{\gamma}\beta_{\mathrm{P}} + \gamma\beta_{\mathrm{D}} \leq \frac{b(a^2\gamma+1)}{\gamma(a^2+b^2)}.$$

Consequently, for the upper bound on λ implied by Theorem 3.3, denoted by $\lambda_{\max}$, it holds that

$$\lambda_{\max} := 2\left(1 + \frac{1}{\gamma}\beta_{\mathrm{P}} + \gamma\beta_{\mathrm{D}}\right) \leq 2 + 2\frac{b(a^2\gamma+1)}{\gamma(a^2+b^2)}.$$

As this bound matches the upper bound on λ from (6.2), the proof is completed.

♠ 6.1*(iii)*: For example, consider $a = 2, b = -1$, for which the trace of $T_{\mathrm{P}}, T_{\mathrm{D}}$ and T_{PD} are all equal to -2.

♠ Finally, we will discuss the range of relaxation parameters λ covered by Theorem 5.2. To this end, the semimonotonicity parameters $\mu_A, \rho_A, \mu_B, \rho_B$ need to be selected such that the relaxation parameter is maximized, i.e.

$$\lambda_{\max}^{\mathrm{semi}} := \max_{\mu_A, \rho_A, \mu_B, \rho_B} 2\left(1 + \frac{1}{\gamma}(\rho_A \square \rho_B) + \gamma(\mu_A \square \mu_B)\right)$$

$$\text{subject to} \quad \mu_A + a^2\rho_A \leq 0 \quad \text{and} \quad \mu_B + b^2\rho_B - b \leq 0,$$

where the constraints are obtained using Proposition 4.5 by enforcing (μ_A, ρ_A)- and (μ_B, ρ_B)-semimonotonicity of operators A and B, respectively. Note that this optimization problem can be encoded as a convex SDP. Solving this problem numerically for $a = 10, b = -1$, the result visualized in Figure 10 is obtained. Hence, when characterizing the nonmonotonicity of the problem through the semimonotonicity properties of A and B separately rather than through the (oblique) weak Minty condition for the combined primal-dual operator, some looseness is inevitably introduced. $\qquad\square$
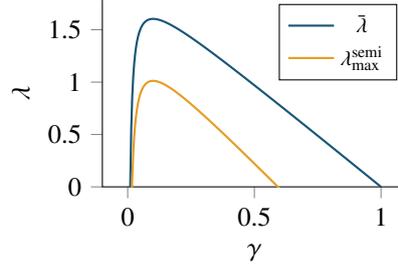
Figure 10: *Comparison between the upper bounds on the relaxation parameter λ in function of the stepsize γ for Example 6.1 with $a = 10$, $b = -1$. $\bar{\lambda}$ is the bound given by (6.2) and $\lambda_{\max}^{\text{semi}}$ is the bound implied by Theorem 5.2.*

Details of Example 6.2 *(nonsmooth optimization).* The corresponding limiting subdifferentials are given by

$$\partial f_A(x) = \begin{cases} -x - 3, & \text{if } x < -1, \\ \{-2, -9\}, & \text{if } x = -1, \\ 6x - 3, & \text{if } x \in (-1, 1), \\ \{-4, 3\}, & \text{if } x = 1, \\ -x - 3, & \text{if } x > 1, \end{cases} \quad \text{and} \quad \partial f_B(x) = \begin{cases} 4x - 1, & \text{if } x < -1, \\ [-5, 1] & \text{if } x = -1, \\ 2x + 3, & \text{if } x \in (-1, 2), \\ [7, 19], & \text{if } x = 2, \\ 2x + 15, & \text{if } x > 2. \end{cases} \quad \text{(C.9)}$$

♠ Assumption II.A1: Follows by definition of the limiting subdifferential [72, Def. 8.3(b)].

♠ Assumption II.A2: It is easy to verify that for $\gamma \neq 1$ it holds that $\mathcal{R}(1 + \gamma \partial f_A) = \mathcal{R}(1 + \gamma \partial f_B) = \mathbb{R}$. Hence, the resolvents $J_{\gamma \partial f_A}$ and $J_{\gamma \partial f_B}$ have full domain.

♠ Assumption III: Let $(x^\star, y^\star) = (0, 3)$. Then, it is of immediate verification that the subdifferentials ∂f_A and ∂f_B are semimonotone at respectively $(x^\star, -y^\star)$ and $(x^\star, y^\star)$ with $\mu_A = -1.2$, $\rho_A = 0.2$, $\mu_B = 1.6$ and $\rho_B = 0.1$.

♠ Follows by applying Theorem 5.2 and using that $\mu_A \square \mu_B = -24/5$ and $\rho_A \square \rho_B = 2/30$. □

Details of Example 6.4 *(stationary point).* Define $g_1(x) := (1 - \sqrt{1 - 4\mu_A \rho_A}/2\rho_A)x - 1$, $g_2(x) := (1 + \sqrt{1 - 4\mu_A \rho_A}/2\rho_A)x - 1$, $h_1(x) := 2x + 1$ and $h_2(x) := \frac{1}{2}x + 1$. Then, the limiting subdifferentials of f_A and f_B are given by

$$\partial f_A(x) = \begin{cases} \{g_1(x), g_2(x)\}, & \text{if } x \in \{-3, 3\}, \\ g_1(x), & \text{if } x \in (-3, 3), \\ g_2(x), & \text{otherwise}, \end{cases} \quad \text{and} \quad \partial f_B(x) = \begin{cases} \{h_1(x), h_2(x)\}, & \text{if } x \in \{-1, 1\}, \\ h_1(x), & \text{if } x \in (-1, 1), \\ h_2(x), & \text{otherwise}. \end{cases}$$

♠ Assumption II.A1: Follows by definition of the limiting subdifferential [72, Def. 8.3(b)].

♠ Assumption II.A2: The resolvents $J_{\gamma \partial f_A}$ and $J_{\gamma \partial f_B}$ have full domain iff $\mathcal{R}(1 + \gamma \partial f_A) = \mathcal{R}(1 + \gamma \partial f_B) = \mathbb{R}$. It is easy to verify that $\mathcal{R}(1 + \gamma \partial f_B) = \mathbb{R}$ for all $\gamma \in \mathbb{R}_+$. On the other hand, we have that

$$(1 + \gamma \partial f_A)(x) = \begin{cases} x - \gamma + \left\{ \frac{1 - \sqrt{1 - 4\mu_A \rho_A}}{2\rho_A}\gamma x, \frac{1 + \sqrt{1 - 4\mu_A \rho_A}}{2\rho_A}\gamma x \right\}, & \text{if } x \in \{-3, 3\}, \\ x - \gamma + \frac{1 - \sqrt{1 - 4\mu_A \rho_A}}{2\rho_A}\gamma x, & \text{if } x \in (-3, 3), \\ x - \gamma + \frac{1 + \sqrt{1 - 4\mu_A \rho_A}}{2\rho_A}\gamma x, & \text{otherwise} \end{cases}$$

from which follows that $\mathcal{R}(1 + \gamma \partial f_A) = \mathbb{R}$ for all $\gamma \in \mathbb{R}_+$ iff γ is selected according to the provided range.

♠ Assumption III: Let $(x^\star, y^\star) = (0, -1)$. Then, it is of immediate verification that the subdifferentials ∂f_A and ∂f_B are semimonotone at respectively $(x^\star, -y^\star)$ and $(x^\star, y^\star)$ for the given parameters $\mu_A, \rho_A, \mu_B, \rho_B$. Furthermore, these parameters satisfy Assumption III.A1.

♠ Follows by applying Theorem 5.2 and using that $\mu_A \square \mu_B = -6/5$ and $\rho_A \square \rho_B = -4/30$. □

References

[1] A. ALACAOGLU, A. BÖHM, AND Y. MALITSKY, *Beyond the golden ratio for variational inequality algorithms*, Journal of Machine Learning Research, 24 (2023), pp. 1–33.

[2] A. ALACAOGLU, D. KIM, AND S. J. WRIGHT, *Extending the reach of first-order algorithms for nonconvex min-max problems with cohypomonotonicity*, 2024, arXiv preprint arXiv:2402.05071.

[3] E. ALLEVI, A. GNUDI, AND I. V. KONNOV, *The proximal point method for nonmonotone variational inequalities*, Mathematical Methods of Operations Research, 63 (2006), pp. 553–565.

[4] F. J. ARAGÓN ARTACHO AND M. H. GEOFFROY, *Characterization of metric regularity of subdifferentials*, Journal of Convex Analysis, 15 (2008), pp. 365–380.

[5] F. J. ARAGÓN ARTACHO, A. DONTCHEV, AND M. GEOFFROY, *Convergence of the proximal point method for metrically regular mappings*, ESAIM Proceedings, 17 (2007), pp. 1–8, https://doi.org/10.1051/proc:071701.

[6] H. ATTOUCH AND M. THÉRA, *A general duality principle for the sum of two operators*, Journal of Convex Analysis, 3 (1996), pp. 1–24.

[7] A. AUSLENDER AND M. TEBOULLE, *Asymptotic cones and functions in optimization and variational inequalities*, Springer Science & Business Media, Berlin, 2006.

[8] S. BARTZ, M. N. DAO, AND H. M. PHAN, *Conical averagedness and convergence analysis of fixed point algorithms*, Journal of Global Optimization, 82 (2022), pp. 351–373.

[9] H. H. BAUSCHKE AND P. L. COMBETTES, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer International Publishing, New York, second ed., 2017.

[10] H. H. BAUSCHKE, W. M. MOURSI, AND X. WANG, *Generalized monotone operators and their averaged resolvents*, Mathematical Programming, 189 (2021), pp. 55–74.

[11] D. S. BERNSTEIN, *Matrix Mathematics: Theory, Facts, and Formulas*, Princeton University Press, Princeton, second ed., 2009.

[12] A. BÖHM, *Solving nonconvex-nonconcave min-max problems exhibiting weak Minty solutions*, 2022, arXiv preprint arXiv:2201.12247.

[13] K. BREDIES, E. CHENCHENE, D. A. LORENZ, AND E. NALDI, *Degenerate preconditioned proximal point algorithms*, SIAM Journal on Optimization, 32 (2022), pp. 2376–2401.

[14] H. BRÉZIS AND P. L. LIONS, *Produits infinis de résolvantes*, Israel Journal of Mathematics, 29 (1978), pp. 329–345.

[15] M. N. BUI AND P. L. COMBETTES, *Warped proximal iterations for monotone inclusions*, Journal of Mathematical Analysis and Applications, 491 (2020), p. 124315.

[16] R. S. BURACHIK AND A. N. IUSEM, *Set-Valued Mappings and Enlargements of Monotone Operators*, Springer Optimization and Its Applications, Springer, June 2008.

[17] J. V. BURKE AND M. QIAN, *A variable metric proximal point algorithm for monotone operators*, SIAM Journal on Control and Optimization, 37 (1999), pp. 353–375.

[18] 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 (2011), pp. 120–145.

[19] P. L. COMBETTES AND T. PENNANEN, *Proximal methods for cohypomonotone operators*, SIAM Journal on Control and Optimization, 43 (2004), pp. 731–742.

[20] L. CONDAT, *A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms*, Journal of Optimization Theory and Applications, 158 (2013), pp. 460–479.

[21] M. N. Dao and H. M. Phan, *Adaptive Douglas–Rachford splitting algorithm for the sum of two operators*, SIAM Journal on Optimization, 29 (2019), pp. 2697–2724.

[22] C. Daskalakis, D. J. Foster, and N. Golowich, *Independent policy gradient methods for competitive reinforcement learning*, Advances in neural information processing systems, 33 (2020), pp. 5527–5540.

[23] D. Davis and W. Yin, *Convergence rate analysis of several splitting schemes*, 2015, arXiv preprint arXiv:1406.4834.

[24] J. Diakonikolas, C. Daskalakis, and M. I. Jordan, *Efficient methods for structured nonconvex-nonconcave min-max optimization*, in International Conference on Artificial Intelligence and Statistics, 2021, pp. 2746–2754.

[25] A. L. Dontchev and R. T. Rockafellar, *Implicit functions and solution mappings*, Springer Monographs in Mathematics. Springer, 208 (2009).

[26] 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 (1956), pp. 421–439.

[27] D. Drusvyatskiy and A. S. Lewis, *Error bounds, quadratic growth, and linear convergence of proximal methods*, Mathematics of Operations Research, 43 (2018), pp. 919–948.

[28] J. Eckstein, *The Lions–Mercier splitting algorithm and the alternating direction method are instances of the proximal point algorithm*, tech. report, LIDS-P-1769, 1988.

[29] J. Eckstein, *Splitting Methods for Monotone Operators with Applications to Parallel Optimization*, PhD thesis, Massachusetts Institute of Technology, 1989.

[30] J. Eckstein, *Nonlinear proximal point algorithms using Bregman functions, with applications to convex programming*, Mathematics of Operations Research, 18 (1993), pp. 202–226.

[31] J. Eckstein and D. P. Bertsekas, *On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators*, Mathematical Programming, 55 (1992), pp. 293–318.

[32] E. Esser, X. Zhang, and T. F. Chan, *A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 1015–1046.

[33] F. Facchinei and J.-S. Pang, *Finite-Dimensional Variational Inequalities and Complementarity Problems, Volume I*, Springer Series in Operations Research, Springer, New York, 2003.

[34] N. Farouq, *Pseudomonotone variational inequalities: Convergence of proximal methods*, Journal of Optimization Theory and Applications, 109 (2001), pp. 311–326.

[35] F. Giannessi, *On Minty variational principle*, in New Trends in Mathematical Programming: Homage to Steven Vajda, Springer US, Boston, MA, 1998, pp. 93–99.

[36] P. Giselsson, *Nonlinear forward-backward splitting with projection correction*, SIAM Journal on Optimization, 31 (2021), pp. 2199–2226.

[37] P. Giselsson and S. Boyd, *Linear convergence and metric selection for Douglas-Rachford splitting and ADMM*, IEEE Transactions on Automatic Control, 62 (2017), pp. 532–544.

[38] E. Gorbunov, N. Loizou, and G. Gidel, *Extragradient method: $\mathcal{O}(1/k)$ last-iterate convergence for monotone variational inequalities and connections with cocoercivity*, in International Conference on Artificial Intelligence and Statistics, 2022, pp. 366–402.

[39] E. Gorbunov, A. Taylor, S. Horváth, and G. Gidel, *Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity*, in International Conference on Machine Learning, PMLR, 2023, pp. 11614–11641.

[40] B. Grimmer, H. Lu, P. Worah, and V. Mirrokni, *The landscape of the proximal point method for nonconvex–nonconcave minimax optimization*, Mathematical Programming, (2022), pp. 1–35.

[41] K. Guo, D. Han, and X. Yuan, *Convergence analysis of Douglas–Rachford splitting method for "strongly + weakly" convex programming*, SIAM Journal on Numerical Analysis, 55 (2017), pp. 1549–1577.

[42] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 119–149.

[43] J.-B. Hiriart-Urruty and C. Lemarechal, *Convex analysis and minimization algorithms I : fundamentals.*, Die Grundlehren der mathematischen Wissenschaften 305, Springer, Berlin, 2nd corr. ed. ed., 1996.

[44] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge university press, Cambridge, 2012.

[45] Z. Huang and M. A. Noor, *An explicit projection method for a system of nonlinear variational inequalities with different (γ, r)-cocoercive mappings*, Applied Mathematics and Computation, 190 (2007), pp. 356–361.

[46] A. N. Iusem, T. Pennanen, and B. F. Svaiter, *Inexact variants of the proximal point algorithm without monotonicity*, SIAM Journal on Optimization, 13 (2003), pp. 1080–1097.

[47] I. V. Konnov, *A class of combined iterative methods for solving variational inequalities*, Journal of Optimization Theory and Applications, 94 (1997), pp. 677–693.

[48] P. Latafat, N. M. Freris, and P. Patrinos, *A new randomized block-coordinate primal-dual proximal algorithm for distributed optimization*, IEEE Transactions on Automatic Control, 64 (2019), pp. 4050–4065.

[49] P. Latafat and P. Patrinos, *Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators*, Computational Optimization and Applications, 68 (2017), pp. 57–93.

[50] S. Lee and D. Kim, *Fast extra gradient methods for smooth structured nonconvex-nonconcave minimax problems*, Advances in Neural Information Processing Systems, 34 (2021), pp. 22588–22600.

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

[52] G. Li, T. Liu, and T. K. Pong, *Peaceman–Rachford splitting for a class of nonconvex optimization problems*, Computational Optimization and Applications, 68 (2017), pp. 407–436.

[53] G. Li and T. K. Pong, *Douglas–Rachford splitting for nonconvex optimization with application to nonconvex feasibility problems*, Mathematical Programming, 159 (2016), pp. 371–401.

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

[55] J. Luque, *A nonlinear proximal point algorithm*, in 26th IEEE Conference on Decision and Control, vol. 26, 1987, pp. 816–817.

[56] B. Martinet, *Régularisation d'inéquations variationnelles par approximations successives*, Revue française d'informatique et de recherche opérationnelle. Série rouge, 4 (1970), pp. 154–158.

[57] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*, vol. 71, SIAM, Philadelphia, 2000.

[58] G. J. Minty, *Monotone (Nonlinear) Operators in Hilbert space*, Duke Mathematical Journal, 29 (1962), pp. 341 – 346.

[59] R. D. C. Monteiro and C.-K. Sim, *Complexity of the relaxed Peaceman–Rachford splitting method for the sum of two maximal strongly monotone operators*, Computational Optimization and Applications, 70 (2018), pp. 763–790.

[60] Y. Nesterov, *Introductory Lectures on Convex Optimization: A Basic Course*, vol. 87 of Applied Optimization, Springer US, Boston, MA, 2004.

[61] J. v. Neumann, *A model of general economic equilibrium*, The Review of Economic Studies, 13 (1945), pp. 1–9.

[62] J. Nocedal and S. Wright, *Numerical Optimization*, Springer Series in Operations Research and Financial Engineering, Springer New York, New York, NY, 2nd ed. 2006. ed., 2006.

[63] R. G. Otero and A. Iusem, *Regularity results for semimonotone operators*, Computational & Applied Mathematics, 30 (2011), p. 13.

[64] N. Parikh and S. Boyd, *Proximal algorithms*, Foundations and Trends in Optimization, 1 (2014), pp. 127–239.

[65] P. Patrinos, L. Stella, and A. Bemporad, *Douglas–Rachford splitting: Complexity estimates and accelerated variants*, in 53rd IEEE Conference on Decision and Control, 2014, pp. 4234–4239.

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

[67] T. Pennanen, *Local convergence of the proximal point algorithm and multiplier methods without monotonicity*, Mathematics of Operations Research, 27 (2002), pp. 170–191.

[68] T. Pethick, O. Fercoq, P. Latafat, P. Patrinos, and V. Cevher, *Solving stochastic weak minty variational inequalities without increasing batch size*, 2023, arXiv preprint arXiv:2302.09029.

[69] T. Pethick, P. Latafat, P. Patrinos, O. Fercoq, and V. Cevher, *Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems*, in International Conference on Learning Representations, 2021.

[70] B. Polyak, *Introduction to Optimization*, Optimization Software, New York, 1987.

[71] G. Pontus and W. M. Moursi, *On compositions of special cases of Lipschitz continuous operators*, Fixed Point Theory and Algorithms for Sciences and Engineering, 2021 (2021).

[72] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, 1970.

[73] R. T. Rockafellar, *Augmented Lagrangians and applications of the proximal point algorithm in convex programming*, Mathematics of Operations Research, 1 (1976), pp. 97–116.

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

[75] R. T. Rockafellar, *Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity*, Set-Valued and Variational Analysis, 27 (2019), pp. 863–893.

[76] R. T. Rockafellar and J. Sun, *Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging*, Mathematical Programming, 174 (2019), pp. 453–471.

[77] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, Springer Science & Business Media, Berlin, 2009.

[78] E. K. Ryu, R. Hannah, and W. Yin, *Scaled relative graphs: Nonexpansive operators via 2D Euclidean geometry*, Mathematical Programming, (2021), pp. 1–51.

[79] L. Shen and S. Pan, *Linear convergence of the generalized PPA and several splitting methods for the composite inclusion problem*, 2016, arXiv preprint arXiv:1508.05156.

[80] M. V. Solodov and B. F. Svaiter, *A hybrid projection-proximal point algorithm.*, Journal of Convex Analysis, 6 (1999), pp. 59–70.

[81] M. V. Solodov and P. Tseng, *Modified projection-type methods for monotone variational inequalities*, SIAM Journal on Control and Optimization, 34 (1996), pp. 1814–1830.

[82] J. E. Spingarn, *Submonotone mappings and the proximal point algorithm*, Numerical Functional Analysis and Optimization, 4 (1982), pp. 123–150.

[83] B. F. Svaiter, *On weak convergence of the Douglas–Rachford method*, SIAM Journal on Control and Optimization, 49 (2011), pp. 280–287.

[84] A. Themelis and P. Patrinos, *Douglas-Rachford splitting and ADMM for nonconvex optimization: Tight convergence results*, SIAM Journal on Optimization, 30 (2020), pp. 149–181.

[85] Q. Tran-Dinh, *Extragradient-type methods with $\mathcal{O}(1/k)$ last-iterate convergence rates for co-hypomonotone inclusions*, 2023, arXiv preprint arXiv:2302.04099.

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

[87] Z. Zhou, P. Mertikopoulos, N. Bambos, S. Boyd, and P. W. Glynn, *Stochastic mirror descent in variationally coherent optimization problems*, Advances in Neural Information Processing Systems, 30 (2017).