Journal of Optimization Theory and Applications (2025) 206:7
https://doi.org/10.1007/s10957-025-02680-x

Convergence of the Chambolle–Pock Algorithm in the Absence of Monotonicity

Brecht Evens[1] · **Puya Latafat**[2] · **Panagiotis Patrinos**[1]

Received: 9 July 2024 / Accepted: 23 March 2025 / Published online: 5 May 2025
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2025

Abstract

The Chambolle–Pock algorithm (CPA), also known as the primal-dual hybrid gradient method, has gained popularity over the last decade due to its success in solving large-scale convex structured problems. This work extends its convergence analysis for problems with varying degrees of (non)monotonicity, quantified through a so-called oblique weak Minty condition on the associated primal-dual operator. Our results reveal novel stepsize and relaxation parameter ranges which do not only depend on the norm of the linear mapping, but also on its other singular values. In particular, in nonmonotone settings, in addition to the classical stepsize conditions, extra bounds on the stepsizes and relaxation parameters are required. On the other hand, in the strongly monotone setting, the relaxation parameter is allowed to exceed the classical upper bound of two. Moreover, we build upon the recently introduced class of semimonotone operators, providing sufficient convergence conditions for CPA when the individual operators are semimonotone. Since this class of operators encompasses traditional operator classes including (hypo)- and co(hypo)-monotone operators, this analysis recovers and extends existing results for CPA. Tightness of the proposed stepsize ranges is demonstrated through several examples.

Communicated by Aris Daniilidis.

This work was supported by: the Research Foundation Flanders (FWO) PhD grant No. 1183822N; postdoctoral grant No. 12Y7622N and research projects G081222N, G033822N, G0A0920N; European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 953348.

✉ Brecht Evens
brecht.evens@kuleuven.be

Puya Latafat
puya.latafat@imtlucca.it

Panagiotis Patrinos
panos.patrinos@kuleuven.be

[1] Department of Electrical Engineering (ESAT-STADIUS), KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium

[2] IMT School for Advanced Studies Lucca, Piazza S. Francesco 19, 55100 Lucca, Italy

Keywords Convex/nonconvex optimization · Monotone/nonmonotone inclusions · Chambolle–Pock algorithm · Primal-dual hybrid gradient · Semimonotone operators

Mathematics Subject Classification 47H04 · 49J52 · 49J53 · 65K15 · 90C26

1 Introduction

This paper considers composite inclusion problems of the form

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

where $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, $B : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ are two (possibly nonmonotone) operators and $L \in \mathbb{R}^{m\times n}$ is a nonzero matrix. Problems of this form emerge naturally in a wide variety of applications in optimization and variational analysis. For instance, in the framework of convex optimization, the first-order optimality condition for minimizing $g(x) + h(Lx)$ is of the form (P-I), with $A = \partial g$ and $B = \partial h$ representing the subdifferentials of proper lsc convex functions g and h.

One of the central algorithms for solving (P-I) is the Chambolle–Pock algorithm (CPA) [12] (also known as the primal-dual hybrid gradient (PDHG) method [18, 25, 59]). Given strictly positive stepsizes $\gamma, \tau > 0$, a sequence of strictly positive relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$ and an initial guess $(x^0, y^0) \in \mathbb{R}^{n+m}$, this algorithm consists of the following iterates.

$$\begin{aligned} \bar{x}^k &\in J_{\gamma A}\big(x^k - \gamma L^\top y^k\big) \\ \bar{y}^k &\in J_{\tau B^{-1}}\big(y^k + \tau L(2\bar{x}^k - x^k)\big) \\ x^{k+1} &= x^k + \lambda_k(\bar{x}^k - x^k) \\ y^{k+1} &= y^k + \lambda_k(\bar{y}^k - y^k) \end{aligned} \tag{CPA}$$

The convergence analysis of CPA in literature largely relies upon an underlying monotonicity assumption. In this work, we identify classes of nonmonotone problems along with corresponding stepsize and relaxation parameter conditions for which CPA remains convergent. To this end, we rely on casting CPA as an instance of the preconditioned proximal point algorithm (PPPA). This connection was previously exploited in [11, 14, 18, 25, 31] in the monotone setting. Many other widely used numerical methods can also be interpreted as special cases of PPPA, see e.g. [14, 16, 17, 46, 49]. In particular, consider the inclusion problem of finding a zero of a set-valued operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, i.e.,

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

Then, given a symmetric positive *semidefinite* preconditioning matrix $P \in \mathbb{R}^{n\times n}$ and a sequence of strictly positive relaxation parameters $(\lambda_k)_{k\in\mathbb{N}}$, the (relaxed) precondi-

tioned proximal point algorithm applied to (G-I) consists of the fixed point iterations

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

The Chambolle–Pock algorithm can be retrieved for certain choice of the preconditioner P and the operator T. Specifically, consider the so-called *primal-dual inclusion*

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

Then, letting $z^k = (x^k, y^k)$ and $\bar{z}^k = (\bar{x}^k, \bar{y}^k)$, CPA is equivalent to applying PPPA to (PD-I), with preconditioner

$$P = \begin{bmatrix} \frac{1}{\gamma}\mathrm{I}_n & -L^\top \\ -L & \frac{1}{\tau}\mathrm{I}_m \end{bmatrix}. \tag{1}$$

As a result of this equivalence, the convergence properties of CPA can be inferred from those of PPPA. In the monotone setting, convergence of PPPA is well understood, not only for positive definite preconditioners [34, 46–48] but also for positive semidefinite ones [31, Thm. 3.4], [11, §2.1]. Analogously, the convergence of CPA for monotone inclusions is relatively well-understood, provided that the stepsizes γ and τ satisfy a certain stepsize condition. The standard assumption in the first works on CPA such as [12, 18, 25] was that the stepsizes γ and τ satisfy $\gamma\tau\|L\|^2 < 1$. This assumption was later relaxed to $\gamma\tau\|L\|^2 \leq 1$ in [14, 31, 39], broadening the scope of the analysis to Douglas-Rachford splitting (DRS), for which $\tau = 1/\gamma$ and $L = \mathrm{I}$. Interestingly, when interpreting CPA as a particular instance of PPPA, the stepsize condition discussed in these works is directly linked to the positive definiteness of the preconditioning matrix P in PPPA. This connection becomes evident by observing that, owing to the Schur complement lemma, P is positive definite under the traditional stepsize condition $\gamma\tau\|L\|^2 < 1$ and positive semidefinite under the relaxed stepsize condition $\gamma\tau\|L\|^2 \leq 1$. Note that the extended stepsize range $\gamma\tau\|L\|^2 \leq 4/3$ was examined in [5, 57], although they rely upon a more involved Lyapunov analysis. Recently, convergence of PPPA in the nonmonotone setting has been considered in [20] under the assumption that T admits a set of oblique weak Minty solutions, defined as follows.

Definition 1.1 (*V-oblique weak Minty solutions* [20]) Let $V \in \mathbb{R}^{n\times n}$ be symmetric. An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to have V-oblique weak Minty solutions at (a nonempty set) $\mathcal{S}^\star \subseteq \operatorname{zer} T$ on a set $\mathcal{U} \subseteq \mathbb{R}^n \times \mathbb{R}^n$ if

$$\langle v, z - z^\star \rangle \geq \mathrm{q}_V(v), \qquad \text{for all } z^\star \in \mathcal{S}^\star, (z, v) \in \operatorname{gph} T \cap \mathcal{U}, \tag{2}$$

where the quadratic form $\mathrm{q}_V(v) := \langle v, Vv\rangle$. Whenever $\mathcal{U} = \mathbb{R}^n \times \mathbb{R}^n$ we refrain from explicitly mentioning the set $\mathcal{U}$, and whenever $V = \rho\mathrm{I}$ for some $\rho \in \mathbb{R}$, we refer to them as ρ-weak Minty solutions.

One key aspect of this assumption is its generality, as V is allowed to be any (possibly indefinite) symmetric matrix. For instance, if V is equal to the zero matrix, (2) reduces to the classic Minty variational inequality (MVI) [21, 36], while if $V = \rho \mathrm{I}$ the so-called weak MVI is retrieved. In literature, weak MVI and the closely related notion of cohypomonotonicity have been employed in the context of the extragradient and the forward-backward-forward method [9, 15, 23, 43, 44], as well as the classic PPA method [13, 23, 28, 41].

Leveraging the results from [20] and the primal-dual connection between CPA and PPPA, the first part of this work will focus on establishing convergence of CPA under the assumption that the primal-dual operator T_{PD} admits a set of V-oblique weak Minty solutions. To account for the inherent structure present within T_{PD}, we impose a specific block diagonal form for $V = \mathrm{blkdiag}(V_{\mathrm{P}}, V_{\mathrm{D}})$, which depends on the fundamental subspaces of L (see (8) and the discussion thereafter). Furthermore, we demonstrate that by restricting our obtained results to the case where $L = \mathrm{I}$ and $\tau = 1/\gamma$, the convergence results for nonmonotone DRS from [20, Sec. 3] are retrieved.

In contrast to the DRS setting, where the convergence results follow in a straightforward manner from those of PPPA (see proof of [20, Thm. 3.3]), convergence results for CPA are more challenging to obtain, not only due to additional stepsize parameter τ, but mainly due to the additional complexity in the algorithm introduced by the matrix L. This difficulty is overcome by considering the singular value decomposition of L and using the corresponding orthonormal basis to carefully decompose the preconditioner P and the oblique weak Minty matrix V (see proof of Theorem 3.1).

In practice, it might be difficult to determine whether the associated primal-dual operator of a given inclusion problem admits V-oblique weak Minty solutions. This issue will be addressed in the second part of this work, where we introduce the class of (M, R)-semimonotone operators and provide several calculus rules for this class, allowing to verify the existence of V-oblique weak Minty solutions based on the semimonotonicity properties of the underlying operators A and B. The class of semimonotone operators is defined as follows.

Definition 1.2 (*semimonotonicity*) Let $M, R \in \mathbb{R}^{n\times n}$ be symmetric (possibly indefinite) matrices. An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be (M, R)-semimonotone at $(x', y') \in \mathrm{gph}\, A$ if

$$\langle x - x', y - y' \rangle \geq \mathrm{q}_M(x - x') + \mathrm{q}_R(y - y'), \qquad \text{for all } (x, y) \in \mathrm{gph}\, A, \tag{3}$$

where $\mathrm{q}_X(\cdot) := \langle \cdot, \cdot \rangle_X$ for any symmetric matrix $X \in \mathbb{R}^{n\times n}$. An operator A is said to be (M, R)-semimonotone if it is (M, R)-semimonotone at all $(x', y') \in \mathrm{gph}\, A$. It is said to be maximally (M, R)-semimonotone if its graph is not strictly contained in the graph of another (M, R)-semimonotone operator.

Throughout, whenever $M = \mu \mathrm{I}_n$ and $R = \rho \mathrm{I}_n$ where $\mu, \rho \in \mathbb{R}$, the prefix (M, R) is replaced by (μ, ρ) and condition (3) reduces to

$$\langle x - x', y - y' \rangle \geq \mu \|x - x'\|^2 + \rho \|y - y'\|^2, \qquad \text{for all } (x, y) \in \mathrm{gph}\, A. \tag{4}$$

The class of (μ, ρ)-semimonotone operators was introduced in [20, Sec. 4] and enjoys a lot of additional freedom compared to more traditional operators classes.

For instance, it encompasses the classes of (hypo)monotone, co(hypomonotone), ρ-semimonotone [40, Def. 2], averaged and firmly nonexpansive operators (see [20, Rem. 4.2 & Fig. 5]).

In this work, this notion is generalized by characterizing the operator class with matrices (M, R) instead of scalars (μ, ρ). This generalization is crucial to capture and exploit the specific structure emerging in CPA. To illustrate this, the next theorem provides a simplified version of our main result (see Corollary 5.2 for the full statement). For instance, if μ_A is positive, $(\mu_A L^\top L, \rho_A \mathrm{I}_n)$-semimonotonicity of A in Theorem 1.1 could be replaced by $(\mu_A \|L\|^2, \rho_A)$-semimonotonicity, which is in general a much more restrictive assumption.

Theorem 1.1 (convergence of CPA under semimonotonicity (simplified)) *Let operators $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ and $B : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be outer semicontinuous. Suppose that there exists $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ such that A is $(\mu_A L^\top L, \rho_A \mathrm{I}_n)$-semimonotone at $(x^\star, -L^\top y^\star) \in \operatorname{gph} A$, B is $(\mu_B \mathrm{I}_m, \rho_B L L^\top)$-semimonotone at $(Lx^\star, y^\star) \in \operatorname{gph} B$ and the moduli $(\mu_A, \mu_B, \rho_A, \rho_B) \in \mathbb{R}^4$ satisfy one of the following conditions.*

(i) *(either) $\mu_A = \mu_B = 0$ and $\rho_A = \rho_B = 0$.*
(ii) *(or) $\mu_A + \mu_B > 0$ and $\rho_A = \rho_B = 0$.*
(iii) *(or) $\rho_A + \rho_B > 0$ and $\mu_A = \mu_B = 0$.*
(iv) *(or) $\mu_A + \mu_B > 0$, $\rho_A + \rho_B > 0$ and $\min\{0, \frac{\mu_A \mu_B}{\mu_A + \mu_B}\} \min\{0, \frac{\rho_A \rho_B}{\rho_A + \rho_B}\} < \frac{1}{4\|L\|^2}$.*

Then, there exist positive stepsizes γ, τ and relaxation sequences $(\lambda_k)_{k\in\mathbb{N}}$ such that if the resolvents $J_{\gamma A}, J_{\tau B^{-1}}$ have full domain,[1] any sequence $(\bar{z}^k)_{k\in\mathbb{N}} = (\bar{x}^k, \bar{y}^k)_{k\in\mathbb{N}}$ generated by CPA starting from $z^0 \in \mathbb{R}^n$ either reaches a point $\bar{z}^k \in \operatorname{zer} T_{\mathrm{PD}}$ in a finite number of iterations or every limit point of $(\bar{z}^k)_{k\in\mathbb{N}}$ belongs to $\operatorname{zer} T_{\mathrm{PD}}$.

This convergence result possesses two primary attributes that deserve attention. First of all, it only requires semimonotonicity of the involved operators at a single point, as opposed to the traditional global assumptions of (hypo)monotonicity and co(hypo)monotonicity. Secondly, by considering the more general class of semimonotone operators, we obtain fundamentally new convergence results (see case (iv)), not covered by any existing theory for CPA. Most notably, this includes examples where $\mu_A \mu_B < 0$ and $\rho_A \rho_B < 0$, for which neither the primal nor the dual nor the primal-dual inclusion are monotone (see e.g. Example 5.3).

As first observed in [51], CPA can be viewed as a particular instance of proximal ADMM. Exploiting this connection it is possible to obtain convergence results for CPA based on those for nonconvex proximal ADMM, see [8, 32]. This approach leads to requirements for L such as a full row rank assumption and restrictions on its condition number (see [8, Ass. 1, Rem. 2(c)]). Our convergence results for CPA also do not depend on any explicit rank conditions on L, allowing to cover rank-deficient cases without introducing a switching mechanism.

[1] The full domain assumption is imposed to ensure that the iterates of CPA are well-defined.

1.1 Contributions

The main contribution of the paper is to establish convergence of CPA under the assumption that the primal-dual operator T_{PD} admits a set of V-oblique weak Minty solutions, which leads to novel stepsize and relaxation parameter ranges in both strongly monotone and nonmonotone settings (see Theorem 3.1 and the preceding discussion). Interestingly, in contrast to the classical stepsize condition $\gamma\tau\|L\|^2 \leq 1$ in the monotone setting, the conditions obtained through our analysis not only depend on the norm of L but also on its other singular values. The tightness of our main convergence theorem is demonstrated through Examples 3.1 and 3.2.

As our second main contribution, convergence results are provided for the class of semimonotone operators [20, Sec. 4], which can be viewed as a natural extension of the (hypo)- and co(hypo)monotone operators. We show that the stepsize requirements reduce to a look-up table depending on the level of (hypo)- and co(hypo)monotonicity (see Corollary 5.2). These results are made possible by establishing a link between the oblique weak Minty assumption for the primal-dual operator and semimonotonicity of the underlying operators A and B, relying on the extended calculus rules developed in Sect. 4 (see also Theorem 5.1).

1.2 Organization

The paper is structured in the following manner. In Sect. 1.3, some notation and standard definitions are provided. Section 2 recalls the main convergence results from [20] for PPPA in the nonmonotone setting. In Sect. 3, the primal-dual equivalence between CPA and PPPA is established, which lead to convergence of CPA under an oblique weak Minty assumption on the associated primal-dual operator. Section 4 discusses and introduces various calculus rules for the class of (M, R)-semimonotone operators. Leveraging these calculus rules, Sect. 5 presents a set of sufficient conditions for the convergence of CPA, based on the semimonotonicity of the underlying operators, along with several examples. Finally, Sect. 6 concludes the paper. 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 use the notation $(w^k)_{k\in I}$ 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 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. The identity matrix is denoted by $\mathrm{I}_n \in \mathbb{R}^{n\times n}$ and the zero matrix by $0_{m\times n} \in \mathbb{R}^{m\times n}$; we write respectively I and 0 when no ambiguity occurs. Adopting the notation from [7], we say a matrix $P \in \mathbb{R}^{m\times n}$ is empty if $\min(m, n) = 0$ and use the conventions $P0_{n\times 0} = 0_{m\times 0}$, $0_{0\times m}P = 0_{0\times n}$ and $0_{m\times 0}0_{0\times n} = 0_{m\times n}$. Moreover, we use the

convention that $\frac{1}{0} = \infty$ and that $0 \cdot \infty = 0$. Given a matrix $P \in \mathbb{R}^{m\times n}$, we denote the range of P by $\mathcal{R}(P)$ and the kernel of P by $\mathcal{N}(P)$. The trace of a square matrix $P \in \mathbb{R}^{n\times n}$ is denoted by tr P.

We denote by $\mathbb{R}^n$ the standard n-dimensional Euclidean space with inner product $\langle\cdot,\cdot\rangle$ and induced norm $\|\cdot\|$. The set of symmetric n-by-n matrices is denoted by $\mathbb{S}^n$. Given a symmetric matrix $P \in \mathbb{S}^n$, we write $P \succeq 0$ and $P \succ 0$ to denote that P is positive semidefinite and positive definite, respectively. Furthermore, for any $P \in \mathbb{S}^n$ we define the quadratic form $\mathrm{q}_P(x) := \langle x, Px\rangle$. Let diag$(\cdot)$ denote the diagonal matrix whose arguments constitute its diagonal elements. For arbitrary matrices A and B, we define the direct sum $A \oplus B = \mathrm{blkdiag}(A, B)$, where blkdiag$(\cdot)$ denotes the block diagonal matrix whose arguments constitute its diagonal blocks. We denote the Kronecker product between two matrices of arbitrary size by $\otimes$.

Two vectors $u, v \in \mathbb{R}^n$ are said to be orthogonal if $\langle u, v\rangle = 0$, and orthonormal if they are orthogonal and $\|u\| = \|v\| = 1$. Two linear subspaces $\mathbf{U} \subseteq \mathbb{R}^n$ and $\mathbf{V} \subseteq \mathbb{R}^n$ are said to be orthogonal if any $u \in \mathbf{U}$ and any $v \in \mathbf{V}$ are orthogonal. We say that $U \in \mathbb{R}^{n\times m}$ is an orthonormal basis for a linear subspace $\mathbf{U} \subseteq \mathbb{R}^n$ if U has orthonormal columns and $\mathcal{R}(U) = \mathbf{U}$.

The effective domain of an extended-real-valued function $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ is given by the set $\mathrm{dom}\, f := \{x \in \mathbb{R}^n \,|\, f(x) < \infty\}$. We say that f is proper if $\mathrm{dom}\, f \neq \emptyset$ and that f is lower semicontinuous (lsc) if the epigraph $\mathrm{epi}\, f := \{(x,\alpha) \in \mathbb{R}^n \times \mathbb{R} \,|\, f(x) \leq \alpha\}$ is a closed subset of $\mathbb{R}^{n+1}$. We denote the limiting subdifferential of f by ∂f. We denote the normal cone of a set $E \subseteq \mathbb{R}^n$ by N_E and the projection onto E is denoted by $\Pi_E(x) := \arg\min_{z\in E} \|z - x\|$. 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 $\mathrm{dom}\, A := \{x \in \mathbb{R}^n \mid Ax \neq \emptyset\}$, its graph by $\mathrm{gph}\, A := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^d \mid y \in Ax\}$, and the set of its zeros by $\mathrm{zer}\, A := \{x \in \mathbb{R}^n \mid 0 \in Ax\}$. The inverse of A is defined through its graph: $\mathrm{gph}\, A^{-1} := \{(y, x) \mid (x, y) \in \mathrm{gph}\, A\}$. The *resolvent* of A is defined by $J_A := (\mathrm{id}+A)^{-1}$. We say that A is *outer semicontinuous (osc)* at $x' \in \mathrm{dom}\, A$ if $\limsup_{x\to x'} Ax := \{y \mid \exists x^k \to x', \exists y^k \to y \text{ with } y^k \in Ax^k\} \subseteq Ax'$. A being osc everywhere is equivalent to its graph being a closed subset of $\mathbb{R}^n \times \mathbb{R}^d$. An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be monotone if $\langle x - x', y - y'\rangle \geq 0$ for all $(x, y), (x', y') \in \mathrm{gph}\, A$. Furthermore, it is said to be μ-monotone for some $\mu \in \mathbb{R}$ if $A - \mu\mathrm{id}$ is monotone and ρ-comonotone for some $\rho \in \mathbb{R}$ if $A^{-1} - \rho\mathrm{id}$ is monotone.

2 Preliminaries on the Preconditioned Proximal Point Algorithm

Departing from the classical monotone setting, convergence of relaxed PPPA was established in [20] for a class of nonmonotone operators that admit a set of *oblique weak Minty solutions* (see Definition 1.1). This result will serve as our primary tool for establishing convergence of CPA in the nonmonotone setting, which is why we will reiterate it here. In particular, its analysis involves the following assumptions.

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

A1 $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is outer semicontinuous.
A2 The preconditioned resolvent $(P + T)^{-1} \circ P$ 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 $\mathcal{S}^\star$.
A4 $P \in \mathbb{S}^n$ is a symmetric positive semidefinite matrix such that

$$1 + \bar{\eta} > 0 \quad \text{with} \quad \bar{\eta} := \lambda_{\min}(U^\top V P U), \tag{5}$$

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

As observed in [20, Rem. 2.2], Assumption I.A4 holds if and only if $(U^\top P U)^{-1} + U^\top V U \succ 0$, which is for instance trivially satisfied when V is positive semidefinite. In contrast to the convergence analysis techniques relying on firm nonexpansiveness of the resolvent mapping, the analysis of [20] relies on a projective interpretation of the preconditioned proximal point algorithm, which dates back to [30, 52, 53]. Most notably, it was demonstrated in [20, Lem. 2.3] that the update rule for the (shadow) sequence generated by PPPA can be interpreted as a relaxed projection onto a certain halfspace, and that if any iterate belongs to this halfspace, which contains the set of projected oblique weak Minty solutions $\Pi_{\mathcal{R}(P)}\mathcal{S}^\star$, this implies its optimality. Based on this insight, the following convergence result for PPPA was established.

Theorem 2.1 ([20, Thm. 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 $\lambda_k \in \big(0, 2(1+\bar{\eta})\big)$ such that $\liminf_{k\to\infty} \lambda_k\big(2(1+\bar{\eta}) - \lambda_k\big) > 0$, where $\bar{\eta}$ is defined as in (5). 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 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 in Assumption I.A3 the set $\Pi_{\mathcal{R}(P)}\mathcal{S}^\star$ is equal to $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$, then the sequences $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$, $(\Pi_{\mathcal{R}(P)} \bar{z}^k)_{k\in\mathbb{N}}$ converge 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 $z^\star \in \operatorname{zer} T$.*

The statement of Theorem 2.1(iv) has been slightly extended compared to [20, Thm. 2.4], as the convergence of $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ to some element of $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$ also implies the convergence of $(\Pi_{\mathcal{R}(P)} \bar{z}^k)_{k\in\mathbb{N}}$ to the same point, since $(\bar{v}^k)_{k\in\mathbb{N}}$ converges to zero. We also remark that by Theorem 2.1(iii) the boundedness of $(\bar{z}^k)_{k\in\mathbb{N}}$ is already guaranteed if the preconditioned resolvent is locally bounded in the sense of [50, Def. 5.14].

As discussed in [20], Assumption I.A3 can be further relaxed by only requiring T to have V-oblique weak Minty solutions at $\mathcal{S}^\star$ on $\mathcal{U} = (\mathcal{R}((P + T)^{-1} P) \times \mathcal{R}(P))$ instead. Under this relaxed assumption, all results from Theorem 2.1 remain valid, as the proof of Theorem 2.1 only involves invoking (2) at points in this restricted set. This relaxation will prove to be relevant in Example 3.1.

3 Chambolle–Pock Under Oblique Weak Minty

In the monotone setting, it is well-known that CPA can be interpreted as applying PPPA to the primal-dual operator T_{PD} [11, 14, 18, 25, 31]. Relying upon the abstract duality framework from [1], [2, Sec. 6.9], this equivalence can be extended to the nonmonotone setting. Within this framework, inclusion problems (P-I) and (PD-I) are labelled as the *primal* and the *primal-dual inclusion*, respectively. Related to these two inclusions is the *dual inclusion*, given by

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

A fundamental equivalence property for these inclusions is summarized below.

Proposition 3.1 ([2, Prop. 6.9.2]) *Let $(x, y) \in \mathbb{R}^n \times \mathbb{R}^m$. The following statements are equivalent: (i) $(x, y) \in \operatorname{zer} T_{\mathrm{PD}}$; (ii) $x \in \operatorname{zer} T_{\mathrm{P}}$ and $y \in \operatorname{zer} T_{\mathrm{D}}$; (iii) $(x, -L^\top y) \in \operatorname{gph} A$ and $(Lx, y) \in \operatorname{gph} B$. Furthermore, it holds that $\operatorname{zer} T_{\mathrm{P}} = \{x \mid \exists y : (x, y) \in \operatorname{zer} T_{\mathrm{PD}}\}$ and $\operatorname{zer} T_{\mathrm{D}} = \{y \mid \exists x : (x, y) \in \operatorname{zer} T_{\mathrm{PD}}\}$.*

A solution of the primal inclusion (P-I) (and of the dual inclusion (D-I)) can thus be obtained by finding a solution of the associated primal-dual inclusion. Now, consider applying PPPA to the primal-dual inclusion (PD-I), with the preconditioner P given by (1). Then, each iteration corresponds to first finding $\bar{x}^k$ and $\bar{y}^k$ satisfying the inclusions $\frac{1}{\gamma}x^k - L^\top y^k \in (\frac{1}{\gamma}\mathrm{id} + A)\bar{x}^k$ and $\frac{1}{\tau}y^k - Lx^k \in -2L\bar{x}^k + (\frac{1}{\tau}\mathrm{id} + B^{-1})\bar{y}^k$ and then performing a relaxation step $z^{k+1} = z^k + \lambda_k(\bar{z}^k - z^k)$. Multiplying the two relations by γ and τ, respectively, and reordering the terms, the update rule for CPA is retrieved. This result is summarized in the following lemma.

Lemma 3.1 (equivalence of CPA and PPPA) *Let $z^0 = (x^0, y^0) \in \mathbb{R}^{n+m}$ be the initial guess for CPA and for PPPA applied to the primal-dual inclusion (PD-I), with the preconditioner P given by (1). Then, the 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 CPA satisfy update rule PPPA.*

Leveraging this connection, we will establish the convergence of CPA based on Theorem 2.1 for PPPA. In contrast to the classical stepsize condition $\gamma\tau\|L\|^2 \leq 1$ in the monotone setting, our upcoming analysis will demonstrate that the stepsize condition on γ and τ for CPA in general does not only depend on $\|L\|$, i.e., the largest singular value of L, but also its other singular values. Therefore, let r denote the rank of L, and without loss of generality, let $\sigma_1, \ldots, \sigma_d$ denote its distinct strictly positive singular values in descending order with respective multiplicities $m_1, \ldots, m_d$. Then, it holds that $r = \sum_{i=1}^d m_i$. Define $\Sigma = \sigma_1 \mathrm{I}_{m_1} \oplus \cdots \oplus \sigma_d \mathrm{I}_{m_d} \in \mathbb{R}^{r\times r}$ and consider the singular value decomposition

$$L = \begin{bmatrix} Y & Y' \end{bmatrix} \Big(\Sigma \oplus 0 \Big) \begin{bmatrix} X & X' \end{bmatrix}^\top, \quad Y = \begin{bmatrix} Y_1 & \cdots & Y_d \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & \cdots & X_d \end{bmatrix}, \tag{6}$$

where the zero matrix is in $\mathbb{R}^{(m-r)\times(n-r)}$, $Y_i \in \mathbb{R}^{m\times m_i}$ and $X_i \in \mathbb{R}^{n\times m_i}$, $i \in [d]$, have orthonormal columns that span the eigenspace corresponding to eigenvalue σ_i^2 of $LL^\top$

and $L^\top L$, respectively, and $Y' \in \mathbb{R}^{m\times(m-r)}$ and $X' \in \mathbb{R}^{n\times(n-r)}$ have orthonormal columns which span the null space of $L^\top$ and L, respectively. The projection onto the range and the kernel of L and $L^\top$ can be expressed as [22, Sec. 2.5.2]

$$\begin{aligned}\Pi_{\mathcal{R}(L)} = YY^\top, \Pi_{\mathcal{N}(L)} = \mathrm{I}_n - XX^\top = X'X'^\top,\\ \Pi_{\mathcal{R}(L^\top)} = XX^\top, \Pi_{\mathcal{N}(L^\top)} = \mathrm{I}_m - YY^\top = Y'Y'^\top.\end{aligned} \tag{7}$$

These projections will play a central role in our upcoming analysis. We will work under the following assumptions on the individual operators A and B and the (nonzero) matrix L.

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

- A1 Operators A and B are outer semicontinuous.
- A2 For the selected positive stepsizes the corresponding resolvents have full domain, i.e., $\operatorname{dom} J_{\gamma A} = \mathbb{R}^n$ and $\operatorname{dom} J_{\tau B^{-1}} = \mathbb{R}^m$.
- A3 The set $\operatorname{zer} T_{\mathrm{PD}}$ is nonempty and there exist parameters $\beta_{\mathrm{P}}, \beta'_{\mathrm{P}}, \beta_{\mathrm{D}}, \beta'_{\mathrm{D}} \in \mathbb{R}$ and a nonempty set $\mathcal{S}^\star \subseteq \operatorname{zer} T_{\mathrm{PD}}$ such that the primal-dual operator T_{PD} has V-oblique weak Minty solutions at $\mathcal{S}^\star$, where

$$V := V_{\mathrm{P}} \oplus V_{\mathrm{D}} = \left(\beta_{\mathrm{P}}\Pi_{\mathcal{R}(L^\top)} + \beta'_{\mathrm{P}}\Pi_{\mathcal{N}(L)}\right) \oplus \left(\beta_{\mathrm{D}}\Pi_{\mathcal{R}(L)} + \beta'_{\mathrm{D}}\Pi_{\mathcal{N}(L^\top)}\right) \in \mathbb{S}^{n+m}. \tag{8}$$

Moreover, defining

$$\delta := 1 - [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-\left(\|L\|^2 - \sigma_d^2\right) = \begin{cases} 1, & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 0,\\ 1 + \beta_{\mathrm{P}}\beta_{\mathrm{D}}\left(\|L\|^2 - \sigma_d^2\right) & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0,\end{cases} \tag{9}$$

the following conditions on β_{P}, β'_{P}, β_{D} and β'_{D} hold.

$$\begin{aligned}\beta'_{\mathrm{P}} \geq \begin{cases} 0, & \text{if } \beta_{\mathrm{P}} \geq 0,\\ \frac{\beta_{\mathrm{P}}}{\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}, & \text{if } \beta_{\mathrm{P}} < 0,\end{cases} \qquad \beta'_{\mathrm{D}} \geq \begin{cases} 0, & \text{if } \beta_{\mathrm{D}} \geq 0,\\ \frac{\beta_{\mathrm{D}}}{\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}, & \text{if } \beta_{\mathrm{D}} < 0.\end{cases}\\ \text{and } [\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \tfrac{1}{4\|L\|^2}.\end{aligned} \tag{10}$$

In Assumption II.A3, the matrix V is defined as in (8) with the understanding that in our convergence analysis we shall set $\beta'_{\mathrm{P}} = \infty$ when L is full column rank and $\beta'_{\mathrm{D}} = \infty$ when L is full row rank, as in these cases V is not affected by their choice. Note also that the conditions on β'_{P} and β'_{D} specified in (10) can be slightly relaxed (see [19, Thm. A.1] of the arXiv preprint). This relaxation, however, results in more complicated stepsize rules for γ and τ, and for this reason is not pursued further.

The imposed block structure on V in (8) is not simply an arbitrary choice, but it aligns perfectly with the inherent structure present within the primal-dual operator itself. To illustrate this, consider the following lemma, which shows that the blocks V_{P}

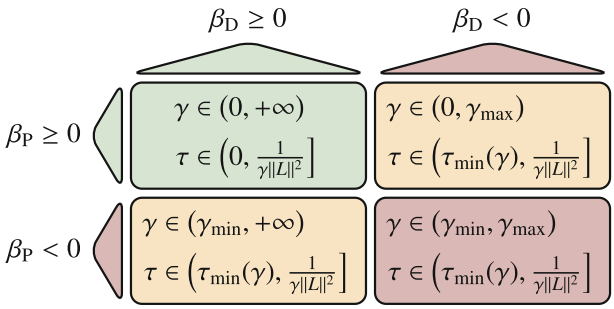

Fig. 1 Range of the stepsizes γ and τ for CPA

and V_D from (8) can be interpreted as the primal and the dual blocks of V, respectively. This lemma extends [20, Lem. 3.2], which considers the case $L = \mathrm{I}$. Notably, the converse of this lemma is not true, as demonstrated in the discussion below [20, Lem. 3.2] through a simple linear example. The proof is deferred to Appendix B.

Lemma 3.2 (oblique weak Minty for primal and dual operator) *Let $V_P \in \mathbb{S}^n$ and $V_D \in \mathbb{S}^m$. Suppose that there exists a nonempty set $\mathcal{S}^\star \subseteq \operatorname{zer} T_{PD}$ such that the primal-dual operator T_{PD} has $(V_P \oplus V_D)$-oblique weak Minty solutions at $\mathcal{S}^\star$ and let $\mathcal{S}_P^\star := \{x^\star \mid \exists y^\star : (x^\star, y^\star) \in \mathcal{S}^\star\} \subseteq \operatorname{zer} T_P$ and $\mathcal{S}_D^\star := \{y^\star \mid \exists x^\star : (x^\star, y^\star) \in \mathcal{S}^\star\} \subseteq \operatorname{zer} T_D$. Then, the primal operator T_P has V_P-oblique weak Minty solutions at $\mathcal{S}_P^\star$ and the dual operator T_D has V_D-oblique weak Minty solutions at $\mathcal{S}_D^\star$.*

One of the main aspects of the upcoming convergence proof for CPA is showing that Assumption I holds for the operator T_{PD} and preconditioner P from (1). Specifically, it will be shown that by decomposing V_P and V_D as in (8), Assumption I.A4 can be split into two distinct conditions, one depending only on β_P and β_D and the other only depending on β_P' and β_D' (see 18). To satisfy both of these conditions, the stepsizes γ and τ and the relaxation parameters λ_k need to be selected as follows.

Stepsize rule I *The stepsizes γ and τ satisfy the bounds provided in Fig. 1, where δ is defined as in (9) and*

$$\gamma_{min} := \frac{-2\beta_P}{\delta + \sqrt{\delta^2 - 4\beta_P\beta_D\|L\|^2}},\ \gamma_{max} := \frac{\delta + \sqrt{\delta^2 - 4\beta_P\beta_D\|L\|^2}}{-2\beta_D\|L\|^2} \text{ and } \tau_{min}(\gamma) := \frac{[-\beta_D(\gamma+\beta_P)]_+}{\gamma(\delta - \beta_P\beta_D\|L\|^2) + \beta_P}. \tag{11}$$

Relaxation parameter rule I *Define*

$$\theta_{\gamma\tau}(\sigma) := \sqrt{\left(\tfrac{1}{2\gamma}\beta_P - \tfrac{1}{2\tau}\beta_D\right)^2 + \beta_P\beta_D\sigma^2}. \tag{12}$$

Let $\bar{\eta} := \min\{\eta, \eta'\}$, where η' is defined as in Fig. 2 and

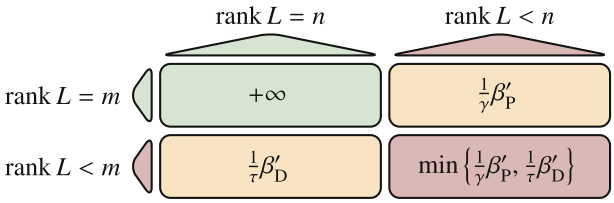

Fig. 2 Definition of η' in Relaxation parameter rule I

$$\eta := \begin{cases} \left\{ \begin{array}{ll} \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0 \\ \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\|L\|), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 0 \end{array} \right\} & \text{if } \gamma\tau < \frac{1}{\|L\|^2}, \\ \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d = 1, \\ \left\{ \begin{array}{ll} \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0 \\ \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_2), & \text{if } \min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0 \\ \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{otherwise} \end{array} \right\} & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d > 1. \end{cases} \tag{13}$$

The sequence $(\lambda_k)_{k\in\mathbb{N}}$ *satisfies* $\lambda_k \in \big(0, 2(1+\bar{\eta})\big)$ *and* $\liminf_{k\to\infty} \lambda_k\big(2(1+\bar{\eta})-\lambda_k\big) > 0$.

When $L = \mathrm{I}$ and $\tau = 1/\gamma$, Stepsize rule I matches the stepsize range for Douglas–Rachford splitting from [20, Thm. 3.3]. In Relaxation parameter rule I there is an interplay between the stepsizes γ and τ and the range of admissible relaxation parameters λ_k. In the monotone setting where $\beta_{\mathrm{P}} = \beta'_{\mathrm{P}} = \beta_{\mathrm{D}} = \beta'_{\mathrm{D}} = 0$, this interplay vanishes as the relaxation rule reduces to the classical condition $\lambda_k \in (0, 2)$. In strongly monotone settings, this interplay allows us to select relaxation parameters beyond the classical upper bound of two. For instance, when $\beta_{\mathrm{P}} > 0$ and $\beta'_{\mathrm{P}} > 0$, then for small enough stepsizes γ the upper bound on λ_k will be larger than two (see [20, Ex. 6.2] for an example in the DRS setting where $L = \mathrm{I}$ and $\tau = 1/\gamma$). Conversely, when $\beta_{\mathrm{D}} > 0$ and $\beta'_{\mathrm{D}} > 0$, this upper bound will be larger than two for small enough τ. Finally, when all β parameters are strictly positive this holds for all valid stepsizes γ and τ (see e.g. Example 3.2).

Having discussed our underlying assumptions and stepsize/relaxation parameter rules, we will now present our main convergence theorem for CPA. The proof relies on carefully decomposing both the preconditioner P and the oblique weak Minty matrix V into two separate, orthogonal matrices. Exploiting the inherent structure present in these orthogonal matrices, the conditions from Assumption I are reduced to a set of eigenvalue problems of two-by-two matrices (see equation (19)).

Theorem 3.1 *Suppose that Assumption II holds, that* γ *and* τ *are selected according to Stepsize rule I and that the relaxation sequence* $(\lambda_k)_{k\in\mathbb{N}}$ *is selected according to Relaxation parameter rule I. Consider the sequences* $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ *and* $(\bar{z}^k)_{k\in\mathbb{N}} = (\bar{x}^k, \bar{y}^k)_{k\in\mathbb{N}}$ *generated by CPA starting from* $z^0 \in \mathbb{R}^{n+m}$. *Then, either a point* $\bar{z}^k \in \operatorname{zer} T_{\mathrm{PD}}$ *is reached in a finite number of iterations or the following hold for the sequences* $(z^k)_{k\in\mathbb{N}}$ *and* $(\bar{z}^k)_{k\in\mathbb{N}}$.

(i) $\bar{v}^k := P(z^k - \bar{z}^k) \in T_{\mathrm{PD}}\bar{z}^k$ *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* $\operatorname{zer} T_{\mathrm{PD}}$.
(iii) *The 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 are in* $\Pi_{\mathcal{R}(P)}\operatorname{zer} T_{\mathrm{PD}}$.
(iv) *If in Assumption II.A3 the set* $\Pi_{\mathcal{R}(P)}\mathcal{S}^\star$ *is equal to* $\Pi_{\mathcal{R}(P)}\operatorname{zer} T$, *then the sequences* $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$, $(\Pi_{\mathcal{R}(P)}\bar{z}^k)_{k\in\mathbb{N}}$ *converge to some element of* $\Pi_{\mathcal{R}(P)}\operatorname{zer} T_{\mathrm{PD}}$. *If additionally* $J_{\gamma A}$ *and* $J_{\tau B^{-1}}$ *are (single-valued) continuous, then* $(\bar{z}^k)_{k\in\mathbb{N}}$ *converges to some* $z^\star \in \operatorname{zer} T_{\mathrm{PD}}$. *Finally, if* λ_k *is additionally uniformly bounded in the interval* $(0, 2)$, *then* $(z^k)_{k\in\mathbb{N}}$ *also converges to* $z^\star \in \operatorname{zer} T_{\mathrm{PD}}$.

Proof Assumption I.A3 is immediate. Outer semicontinuity of T_{PD} follows from that of A and B [50, Theorem 5.7(a)], showing that Assumption I.A1 holds. Since $J_{\gamma A}$ and $J_{\tau B^{-1}}$ have full domain the preconditioned resolvent $(P + T_{\mathrm{PD}})^{-1}P$ has full domain owing to [50, Lemma 12.14], establishing Assumption I.A2. It only remains to show that Assumption I.A4 holds. Before that, we first show that $\bar{\eta}$ as defined in (5) is characterized by the expression given in Relaxation parameter rule I.

Let $Z_i := X_i \oplus Y_i \in \mathbb{R}^{(m+n)\times(2m_i)}$ and $Z' := X' \oplus Y' \in \mathbb{R}^{(m+n)\times(m+n-2r)}$ where Y_i, X_i, Y' and X' are given by (6). Let $Z := [\, z_1 \cdots z_d \,] \in \mathbb{R}^{(m+n)\times(2r)}$, which by construction has orthonormal columns. The preconditioner P in (1) can be decomposed as

$$P \overset{(6)}{=} \begin{bmatrix} \frac{1}{\gamma}\mathrm{I}_n & -X\Sigma Y^\top \\ -Y\Sigma X^\top & \frac{1}{\tau}\mathrm{I}_m \end{bmatrix} \overset{(7)}{=} \underbrace{\begin{bmatrix} \frac{1}{\gamma}XX^\top & -X\Sigma Y^\top \\ -Y\Sigma X^\top & \frac{1}{\tau}YY^\top \end{bmatrix}}_{=: Z\hat{P}Z^\top} + \underbrace{\left(\tfrac{1}{\gamma}X'X'^\top\right) \oplus \left(\tfrac{1}{\tau}Y'Y'^\top\right)}_{=: Z'P'Z'^\top}, \tag{14}$$

where $P' = \frac{1}{\gamma}\mathrm{I}_{n-r} \oplus \frac{1}{\tau}\mathrm{I}_{m-r}$.

$$\hat{P}_i = \begin{bmatrix} \frac{1}{\gamma} & -\sigma_i \\ -\sigma_i & \frac{1}{\tau} \end{bmatrix}, \quad \hat{P} = \left((\hat{P}_1 \otimes \mathrm{I}_{m_1}) \oplus \cdots \oplus (\hat{P}_d \otimes \mathrm{I}_{m_d})\right)$$

Using this decomposition, an orthonormal basis for $\mathcal{R}(P)$ can be constructed. The matrices $\hat{P}_i \succ 0$ for $i = 2, \ldots, d$ since $\gamma\tau\sigma_i^2 < 1$, and $\hat{P}_1 \succeq 0$ since by Stepsize rule I $\gamma\tau\|L\|^2 \leq 1$. Consequently, an orthonormal basis for $\mathcal{R}(\hat{P}_i)$ is given by $\hat{U}_i = \mathrm{I}_2$ for $i = 2, \ldots, d$. For $i = 1$, an orthonormal basis is given by $\hat{U}_1 = \mathrm{I}_2$ if $\gamma\tau\|L\|^2 < 1$ and if $\gamma\tau\|L\|^2 = 1$ by

$$\hat{U}_1 = \sqrt{\tau/(\gamma+\tau)}\begin{bmatrix} 1 \\ -\sqrt{\gamma/\tau} \end{bmatrix}. \tag{15}$$

Owing to the block-diagonal structure of $\hat{P}$, an orthonormal basis for $\mathcal{R}(\hat{P})$ is therefore given by $\hat{U} = (\hat{U}_1 \otimes \mathrm{I}_{m_1}) \oplus \mathrm{I}_{(2r-2m_1)}$. For any $A \in \mathbb{R}^{k\times l}$ and any full row rank $B \in \mathbb{R}^{l\times p}$, it holds that $\mathcal{R}(AB) = \mathcal{R}(A)$ [7, Prop. 2.6.3]. Since Z and Z' are full column rank and P' is full rank, this implies that $\mathcal{R}(Z\hat{P}Z^\top) = \mathcal{R}(Z\hat{U})$ and $\mathcal{R}(Z'P'Z'^\top) = \mathcal{R}(Z')$.

Finally, since $Z\hat{U}$ has orthonormal columns (as $(Z\hat{U})^\top Z\hat{U} = \mathrm{I}$), an orthonormal basis for $\mathcal{R}(P) = \mathcal{R}(Z\hat{P}Z^\top) \oplus \mathcal{R}(Z'\hat{P}'Z'^\top)$ is given by

$$U = \begin{bmatrix} Z\hat{U} & Z' \end{bmatrix} = \begin{bmatrix} Z_1(\hat{U}_1 \otimes \mathrm{I}_{m_1}) & Z_2 & \cdots & Z_d & Z' \end{bmatrix}. \tag{16}$$

Analogously, the V-oblique weak Minty matrix as defined in (8) can be decomposed as

$$V \overset{(7)}{=} \overbrace{\left(\beta_{\mathrm{P}} X X^\top\right) \oplus \left(\beta_{\mathrm{D}} Y Y^\top\right)}^{=: Z\hat{V}Z^\top} + \overbrace{\left(\beta'_{\mathrm{P}} X' X'^\top\right) \oplus \left(\beta'_{\mathrm{D}} Y' Y'^\top\right)}^{=: Z'V'Z'^\top}, \tag{17}$$

where $\hat{V}_i = \mathrm{diag}(\beta_{\mathrm{P}}, \beta_{\mathrm{D}})$, $\hat{V} = \left((\hat{V}_1 \otimes \mathrm{I}_{m_1}) \oplus \cdots \oplus (\hat{V}_d \otimes \mathrm{I}_{m_d})\right)$ and $V' = \beta'_{\mathrm{P}}\mathrm{I}_{n-r} \oplus \beta'_{\mathrm{D}}\mathrm{I}_{m-r}$. Since Z and Z' both have orthonormal columns, i.e., $Z^\top Z = \mathrm{I}_{2r}$ and $Z'^\top Z' = \mathrm{I}_{m+n-2r}$, and since $\mathcal{R}(Z)$ and $\mathcal{R}(Z')$ are orthogonal, it follows from (14)–(17) that

$$\begin{aligned} U^\top V P U &= \begin{bmatrix} \hat{U}^\top Z^\top \\ Z'^\top \end{bmatrix} (Z\hat{V}\hat{P}Z^\top + Z'V'P'Z'^\top) \begin{bmatrix} Z\hat{U} & Z' \end{bmatrix} \\ &= \hat{U}^\top \hat{V}\hat{P}\hat{U} \oplus Z'^\top V' P' Z'. \end{aligned}$$

Therefore, condition (5) becomes

$$1 + \bar{\eta} = 1 + \lambda_{min}(U^\top V P U) = 1 + \min\Big(\overbrace{\lambda_{min}(\hat{U}^\top \hat{V}\hat{P}\hat{U})}^{=:\eta}, \overbrace{\lambda_{min}(Z'^\top V' P' Z')}^{=:\eta'}\Big) > 0. \tag{18}$$

Due to the block diagonal structure of $\hat{U}^\top \hat{V}\hat{P}\hat{U}$, it follows that

$$\begin{aligned} \eta &= \min\left\{\lambda_{\min}(\hat{U}_1^\top \hat{V}_1 \hat{P}_1 \hat{U}_1), \{\lambda_{\min}(\hat{V}_i \hat{P}_i)\}_{i=2}^d\right\} \\ &= \min\{\lambda_{\min}\Big(\hat{U}_1^\top \begin{bmatrix} \frac{1}{\gamma}\beta_{\mathrm{P}} & -\beta_{\mathrm{P}}\sigma_1 \\ -\beta_{\mathrm{D}}\sigma_1 & \frac{1}{\tau}\beta_{\mathrm{D}} \end{bmatrix} \hat{U}_1\Big), \{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\}_{i=2}^d\}, \end{aligned} \tag{19}$$

where $\theta_{\gamma\tau}(\cdot)$ is defined as in (12). Note that $\theta_{\gamma\tau}(\sigma)$ can also be expressed as

$$\theta_{\gamma\tau}(\sigma) := \sqrt{\left(\tfrac{1}{2\gamma}\beta_{\mathrm{P}} - \tfrac{1}{2\tau}\beta_{\mathrm{D}}\right)^2 + \beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma^2} = \sqrt{\left(\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}}\right)^2 - \tfrac{\beta_{\mathrm{P}}\beta_{\mathrm{D}}}{\gamma\tau}(1 - \gamma\tau\sigma^2)}. \tag{20}$$

If $\gamma\tau\|L\|^2 < 1$, then since $\hat{U}_1 = \mathrm{I}_2$ it follows that η is as in (13) since it reduces to

$$\eta = \min\{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\}_{i=1}^{d}$$
$$= \begin{cases} \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0, \\ \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\|L\|), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 0, \end{cases}$$

where equivalence (20) was used in the second equality to show for which singular value the minimum is attained. Otherwise, $\gamma\tau\|L\|^2 = 1$ and we have that $\hat{U}_1 = \sqrt{\tau/(\gamma+\tau)}\begin{bmatrix} 1 \\ -\sqrt{\gamma/\tau} \end{bmatrix}$, which also matches (13) since

$$\eta = \min\{\lambda_{\min}\Big(\tfrac{\tau}{\gamma+\tau}\begin{bmatrix} 1 & -\sqrt{\gamma/\tau} \end{bmatrix}\begin{bmatrix} \frac{1}{\gamma}\beta_{\mathrm{P}} & -\beta_{\mathrm{P}}\sigma_1 \\ -\beta_{\mathrm{D}}\sigma_1 & \frac{1}{\tau}\beta_{\mathrm{D}} \end{bmatrix}\begin{bmatrix} 1 \\ -\sqrt{\gamma/\tau} \end{bmatrix}\Big), \{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\}_{i=2}^{d}\}$$
$$= \min\left\{\tfrac{1}{\gamma}\beta_{\mathrm{P}} + \tfrac{1}{\tau}\beta_{\mathrm{D}}, \{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\}_{i=2}^{d}\right\}$$
$$= \begin{cases} \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{if } d = 1, \\ \left\{ \begin{array}{ll} \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0 \\ \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_2), & \text{if } \min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0 \\ \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{otherwise} \end{array} \right\}, & \text{if } d > 1, \end{cases}$$

where we used that $\sqrt{\gamma\tau}\sigma_1 = 1$ in the first equality and equivalence (20) to show the second equality.

By definition of Z', V' and P', it holds that

$$\eta' = \lambda_{min}\left(\left(\tfrac{1}{\gamma}\beta'_{\mathrm{P}}X'^{\top}X'\right) \oplus \left(\tfrac{1}{\tau}\beta'_{\mathrm{D}}Y'^{\top}Y'\right)\right) = \lambda_{\min}\Big(\left(\tfrac{1}{\gamma}\beta'_{\mathrm{P}}\mathrm{I}_{n-r}\right) \oplus \left(\tfrac{1}{\tau}\beta'_{\mathrm{D}}\mathrm{I}_{m-r}\right)\Big), \tag{21}$$

which matches the definition of η' provided in Fig. 2. Having shown that the values of η and η' match the ones provided in Relaxation parameter rule I, it remains to show that Stepsize rule I ensures that $1 + \bar{\eta} := 1 + \min\{\eta, \eta'\} > 0$ as required by Assumption I.A4. We proceed with two intermediate claims.

Claim 1 *The pair $(\gamma, \tau) \in \mathbb{R}^2_{++}$ satisfies $\gamma\tau \in (0, 1/\|L\|^2]$ and $1 + \eta > 0$ if and only if γ and τ comply with Stepsize rule I.*

Consider the six different cases for η from (13). For the first and fourth case this follows from Lemma A.1(ii), for the second case from Lemmas A.1(i) and A.1(iii) since $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$ by (10) and for the fifth case from Lemma A.1(i). Finally, for the third and the sixth case use Fact 1, by plugging in $\tau = \frac{1}{\gamma\|L\|^2}$ into $1 + \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}} > 0$ and observing that $\delta = 1$.

Claim 2 *If γ and τ comply with Stepsize rule I then $1 + \eta' > 0$.*

Recall that as noted in the discussion after Assumption II, the value of β'_{P} (resp. β'_{D}) is irrelevant when L is full column rank (resp. full row rank), and was considered

equal to ∞ in this case. In view of (21), γ and τ satisfy $1+\eta' > 0$ if and only if

$$\gamma > -\beta'_{\mathrm{P}} \quad \text{and} \quad \tau > -\beta'_{\mathrm{D}}. \tag{22}$$

We distinguish between two cases.

- ♠ $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0$: Then, owing to (10) it holds that $\beta'_{\mathrm{P}} \geq 0$ and $\beta'_{\mathrm{D}} \geq 0$, so that (22) is implied trivially by Stepsize rule I.
- ♠ $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} < 0$: We begin by deriving two auxiliary inequalities that are essential for showing that (22) holds. By definition of δ from (9) and using $\beta_{\mathrm{P}}\beta_{\mathrm{D}} = [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_+ - [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-$ it holds that for any parameter $\kappa \geq 1$ that

$$\delta - \kappa\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2 = 1 + [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-\big(\sigma_d^2 + (\kappa-1)\|L\|^2\big) - \kappa[\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_+\|L\|^2 \geq 1 - \tfrac{\kappa}{4}, \tag{23}$$

where the inequality follows from the fact that either both β_{P} and β_{D} are negative, in which case inequality (10) is used, or only one is negative, in which case the quantity is greater than or equal to 1. Moreover,

$$\begin{aligned} 0 \leq \delta^2 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2 &= \big(\delta - 2\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big)^2 + 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big(\delta - 1 - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big) \\ &= \big(\delta - 2\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big)^2 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big([\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_+\|L\|^2 - [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-\sigma_d^2\big) \\ &\leq \big(\delta - 2\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2\big)^2, \end{aligned} \tag{24}$$

where the last inequality follows from the trivial observation that for any $a \in \mathbb{R}$, $a[a]_+ \geq 0$ and $a[a]_- \leq 0$.

- ◇ $\gamma > -\beta'_{\mathrm{P}}$: Using (24), we obtain that $\delta + \sqrt{\delta^2 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2} \leq 2(\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2)$. Consequently, by the definition of $\gamma_{\min}$ from (11), it follows that

$$\beta_{\mathrm{P}} < 0 \Longrightarrow -\beta'_{\mathrm{P}} \overset{(10)}{=} \tfrac{-\beta_{\mathrm{P}}}{\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2} \leq \gamma_{\min}. \tag{25}$$

Since Stepsize rule I ensures that $\gamma > 0 = \beta'_{\mathrm{P}}$ if $\beta_{\mathrm{P}} \geq 0$ and $\gamma > \gamma_{\min}$ if $\beta_{\mathrm{P}} < 0$, it thus follows from (25) that indeed $\gamma > -\beta'_{\mathrm{P}}$.

- ◇ $\tau > -\beta'_{\mathrm{D}}$: Observe that regardless of the sign of β_{P} it holds that $\gamma(\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2) + \beta_{\mathrm{P}} > 0$; if $\beta_{\mathrm{P}} \geq 0$, then this holds owing to (23) since $\gamma > 0$, if $\beta_{\mathrm{P}} < 0$, then this inequality remains true since in that case $\gamma > \gamma_{\min}$ by Stepsize rule I and using (25). We proceed by first showing that

$$\tau_{\min}(\gamma) = \tfrac{[-\beta_{\mathrm{D}}(\gamma+\beta_{\mathrm{P}})]_+}{\gamma(\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2) + \beta_{\mathrm{P}}} \geq \tfrac{[-\beta_{\mathrm{D}}]_+}{\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}. \tag{26}$$

If $\beta_{\mathrm{D}} \geq 0$, then (26) holds trivially. Let $\beta_{\mathrm{D}} < 0$. If $\beta_{\mathrm{P}} \geq 0$, then $\gamma + \beta_{\mathrm{P}} > 0$ since $\gamma > 0$. If $\beta_{\mathrm{P}} < 0$, then it also holds that $\gamma + \beta_{\mathrm{P}} > 0$ since in that case $\gamma > \gamma_{\min} \geq \frac{-\beta_{\mathrm{P}}}{1 - \beta_{\mathrm{D}}\beta_{\mathrm{P}}\|L\|^2} > -\beta_{\mathrm{P}}$ by Stepsize rule I and using (25). Therefore, it

holds that $-\beta_{\mathrm{D}}(\gamma+\beta_{\mathrm{P}}) > 0$. Having previously established that the denominators in (26) are strictly positive this inequality therefore is equivalent to

$$-\beta_{\mathrm{D}}(\gamma+\beta_{\mathrm{P}})(\delta-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2) \geq -\beta_{\mathrm{D}}(\gamma(\delta-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2)+\beta_{\mathrm{P}}).$$

Substituting δ from (9) and using $\beta_{\mathrm{P}}\beta_{\mathrm{D}} = [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_+ - [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-$, rearrange it to obtain $[\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-^2\sigma_d^2 + [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_+^2\|L\|^2 \geq 0$, which holds trivially. Finally, having established that (26) holds and noting that $\tau > \tau_{\min}(\gamma)$ owing to Stepsize rule I, it follows immediately that $\tau > -\beta'_{\mathrm{D}}$ by definition of β'_{D} from (10).

Combining Claims 1 and 2, it follows that Stepsize rule I is equivalent to the preconditioner P being positive semidefinite and $1+\bar{\eta} = 1+\min\{\eta,\eta'\} > 0$, establishing that Assumption I.A4 holds. Finally, $(P+T_{\mathrm{PD}})^{-1}\circ P$ is (single-valued) continuous if $J_{\gamma A}$ and $J_{\tau B^{-1}}$ are (single-valued) continuous, since $\bar{z}^k \in (P+T_{\mathrm{PD}})^{-1}Pz^k$ if and only if $\bar{x}^k \in J_{\gamma A}\big(x^k - \gamma L^\top y^k\big)$ and $\bar{y}^k \in J_{\tau B^{-1}}\big(y^k + \tau L(2\bar{x}^k - x^k)\big)$. All claims for CPA follow directly from Theorem 2.1, using the equivalence from Lemma 3.1. □

Note that when λ_k is uniformly bounded in the interval $\big(0, 2(1+\bar{\eta})\big)$, a rate of $O(1/N)$ can be obtained for $\min_{k=0,1,\ldots,N}\|\bar{v}^k\|^2$ by telescoping (see [20, thm. 2.4(iv)]).

Observe that Theorem 3.1 discusses not only the convergence of $(z^k)_{k\in\mathbb{N}}$, but also of its projection onto the range of the preconditioner $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$. Notably, convergence of $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ is established under weaker assumptions than for $(z^k)_{k\in\mathbb{N}}$. When P is positive definite, meaning $\gamma\tau < 1/\|L\|^2$, this is irrelevant because in this case the range of P is full. However, in the positive semidefinite case, when $\gamma\tau = 1/\|L\|^2$, these sequences are no longer identitical. This observation is not surprising, as it is a natural extension of the convergence results for DRS, i.e., when $L = \mathrm{I}$ and $\gamma = 1/\tau$. In particular, in the DRS setting it was shown that the convergence of $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ to $\Pi_{\mathcal{R}(P)}\operatorname{zer}T_{\mathrm{PD}}$ is equivalent to the convergence of the sequence $(s^k)_{k\in\mathbb{N}} := (x^k - \gamma y^k)_{k\in\mathbb{N}}$ to a point $s^\star \in \big\{x^\star - \gamma y^\star \,\big|\, (x^\star, y^\star) \in \operatorname{zer}T_{\mathrm{PD}}\big\}$, and a solution of the primal inclusion is obtained through $J_{\gamma A}(s^\star)$ [20, 33, 54]. In the following proposition, this interpretation is generalized to arbitrary L matrices.

Proposition 3.2 (convergent sequences) *Using the SVD of L from (6), define*

$$U = \begin{bmatrix} \sqrt{\frac{\tau}{\gamma+\tau}}X_1 & X_2 & 0 & \cdots & X_d & 0 & X' & 0 \\ -\sqrt{\frac{\gamma}{\gamma+\tau}}Y_1 & 0 & Y_2 & \cdots & 0 & Y_d & 0 & Y' \end{bmatrix}. \tag{27}$$

Consider a sequence $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ *generated by CPA starting from* $z^0 \in \mathbb{R}^{n+m}$, *where* $\gamma\tau = \frac{1}{\|L\|^2}$ *and define*

$$s^k := U^\top\begin{bmatrix} x^k \\ y^k \end{bmatrix} \quad \text{and} \quad \mathcal{T} := \left\{U^\top\begin{bmatrix} x^\star \\ y^\star \end{bmatrix} \,\middle|\, (x^\star, y^\star) \in \operatorname{zer}T_{\mathrm{PD}}\right\}. \tag{28}$$

Then, the limit points of $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ *are in* $\Pi_{\mathcal{R}(P)}\operatorname{zer}T_{\mathrm{PD}}$ *if and only if the limit points of* $(s^k)_{k\in\mathbb{N}}$ *are in* $\mathcal{T}$, *and the sequence* $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ *converges to some element of* $\Pi_{\mathcal{R}(P)}\operatorname{zer}T_{\mathrm{PD}}$ *if and only if* $(s^k)_{k\in\mathbb{N}}$ *converges to some element of* $\mathcal{T}$.

Proof Notice that U is an orthonormal basis for $\mathcal{R}(P)$, obtained by plugging in (15) into (16). As a result, the claims follow directly from (28), using the fact that $\Pi_{\mathcal{R}(P)} = UU^\top$ and that U has orthonormal columns. □

Notably, Proposition 3.2 along with Theorem 3.1 establishes the convergence of an $(m+n-m_1)$-dimensional sequence $(s^k)_{k\in\mathbb{N}}$ of CPA when $\gamma\tau = 1/\|L\|^2$. Since $s^k = 1/\sqrt{1+\gamma^2}(x^k - \gamma y^k)$ when $L = \mathrm{I}$ and $\tau = 1/\gamma$, it follows immediately that Theorem 3.1 matches the convergence results for DRS obtained in [20, Thm. 3.3]. A simple example where $(s^k)_{k\in\mathbb{N}}$ converges while $(z^k)_{k\in\mathbb{N}}$ diverges is provided in Example 3.2.

In the remainder of this section, two theoretical examples are presented to demonstrate some of the main attributes of the convergence results in Theorem 3.1. Supplementary Python code verifying these results can be found on GitHub[2] and the proofs are deferred to Appendix B.

In the first example, the tightness of the bounds on γ, τ and λ_k from Theorem 3.1 is demonstrated through a simple system of linear equations. In this setting, the iterations of CPA can be expressed as a linear dynamical system, so that tight bounds can be obtained by ensuring stability. In this example, an artificial parameter c is introduced when splitting the problem into the form $A + L^\top BL$. While this parameter may appear inconsequential at first sight, it controls the value of β'_{D} within the problem (see Example 3.1(ii)).

Example 3.1 (saddle point problem) Consider the problem of finding a zero of

$$0 \in T_{\mathrm{P}}x = \begin{bmatrix} b\ell^2 & a \\ -a & b\ell^2 \end{bmatrix} x = \overbrace{\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}}^{A} x + \overbrace{\begin{bmatrix} \ell & 0 & 0 \\ 0 & \ell & 0 \end{bmatrix}}^{L^\top} \overbrace{\begin{bmatrix} b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}}^{B} \overbrace{\begin{bmatrix} \ell & 0 \\ 0 & \ell \\ 0 & 0 \end{bmatrix}}^{L} x, \tag{29}$$

where $a, b, \ell \in \mathbb{R}\backslash\{0\}$ and $c \geq 0$. Suppose that $a^2 \neq b^2\ell^4$ when $b < 0$. Any solution to the inclusion problem $0 \in T_{\mathrm{P}}x$ is a minimax solution of $f(x_1, x_2) := ax_1x_2 + \frac{b\ell^2}{2}(x_1^2 - x_2^2)$ when $b > 0$ and a maximin solution when $b < 0$. Consider the sequence $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ generated by applying CPA to (29) with $\tau = \frac{1}{\gamma\|L\|^2}$ and fixed relaxation parameter λ. Then, the following assertions hold.

(i) The sequence $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ converges if and only if $\lambda \in \big(0, \bar{\lambda}\big)$ and $(z^k)_{k\in\mathbb{N}}$ converges if and only if $\lambda \in \big(0, \min\{2, \bar{\lambda}\}\big)$, where

$$\bar{\lambda} := \min\left\{2\Big(1 + \tfrac{b\ell^2}{\gamma(a^2+b^2\ell^4)} + \tfrac{ba^2\ell^2\gamma}{a^2+b^2\ell^4}\Big), 2\left(1+\gamma c\ell^2\right)\right\}. \tag{30}$$

This upper bound is strictly positive if and only if

$$\gamma \in \begin{cases} (0, +\infty), & \text{if } b > 0, \\ \left(\frac{-2b\ell^2}{a^2+b^2\ell^4+|a^2-b^2\ell^4|}, \frac{a^2+b^2\ell^4+|a^2-b^2\ell^4|}{-2ba^2}\right), & \text{if } b < 0. \end{cases} \tag{31}$$

[2] https://github.com/brechtevens/Minty-CP-examples.

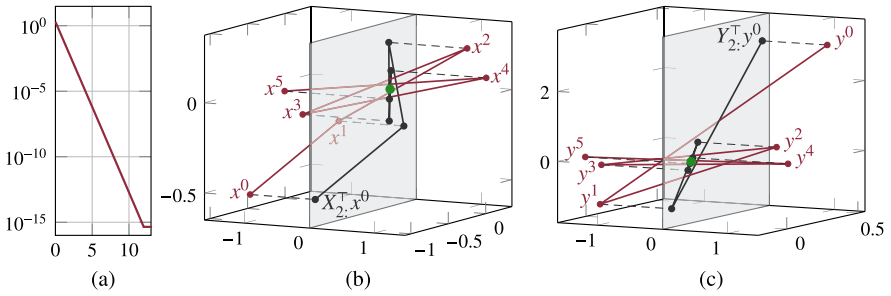

Fig. 3 Convergence of $(s^k)_{k\in\mathbb{N}} = \big(X_1^\top x^k - Y_1^\top y^k,\ X_{2:}^\top x^k,\ Y_{2:}^\top y^k\big)_{k\in\mathbb{N}}$ from Example 3.2 for $n = 3$, $\ell_2 = 1/2$, $\ell_3 = 1/5$ and $\lambda = 2.1$. (a) Norm of $(X_1^\top x^k - Y_1^\top y^k)_{k\in\mathbb{N}}$, which converges to zero. (b) Visualization of the primal sequences $(x^k)_{k\in\mathbb{N}}$ and $(X_{2:}^\top x^k)_{k\in\mathbb{N}}$. Although $(x^k)_{k\in\mathbb{N}}$ does not converge (its first coordinate diverges), its projection onto the 2-dimensional space spanned by the columns of $X_{2:}$ does converge to zero (marked by a green dot). (c) Visualization of the dual sequences $(y^k)_{k\in\mathbb{N}}$ and $(Y_{2:}^\top y^k)_{k\in\mathbb{N}}$. Analogous to the primal setting, $(y^k)_{k\in\mathbb{N}}$ diverges while $(Y_{2:}^\top y^k)_{k\in\mathbb{N}}$ converges to zero

(ii) The operator T_{PD} has a $V = \frac{b\ell^2}{a^2+b^2\ell^4}\Pi_{\mathcal{R}(L^\top)} \oplus \left(\frac{ba^2}{a^2+b^2\ell^4}\Pi_{\mathcal{R}(L)} + c\Pi_{\mathcal{N}(L^\top)}\right)$-oblique weak Minty solution at zero on $\mathcal{U} = \mathbb{R}^5 \times \mathcal{R}(P)$.

(iii) Recalling that Assumption II.A3 only needs to hold on $\mathcal{U}$ (see the discussion below Theorem 2.1), all assumptions of Theorem 3.1(iv) hold. Moreover, this result is tight in the sense that Stepsize rule I corresponds to (31) and the bounds from Relaxation parameter rule I match those from (30).

(iv) The range of parameters a, b, c and ℓ for which CPA converges includes cases where neither the primal, nor the dual, nor the primal-dual problem are monotone. For instance, when $a = 10$, $b = -\frac{1}{4}$, $c = 0$ and $\ell = 2$, by Example 3.1(i), $(z^k)_{k\in\mathbb{N}}$ then converges if and only if $\gamma \in \left(\frac{1}{100}, 1\right)$ and $\lambda \in \left(0, 2 - \frac{2}{101\gamma} - \frac{200\gamma}{101}\right)$.

The second example focusses on a particular instance of Theorem 3.1 where β_{P} and β_{D} are both strictly positive, the number of distinct singular values of L is strictly larger than 1 and $\gamma\tau\|L\|^2 = 1$. Then, the admissible range for λ depends on the second largest singular value of L (see Relaxation parameter rule I). The following example demonstrates that this dependence is not merely an artifact of our analysis, but is also observed in practice. Moreover, Fig. 3 illustrates that using relaxation parameters greater than two may cause the non-projected sequence $(z_k)_{k\in\mathbb{N}}$ to diverge, in line with Theorem 3.1(iv).

Example 3.2 (influence of singular values) Let $n \in \{2, 3\ldots\}$ and $L = \mathrm{diag}(1, \ell_2 \ldots \ell_n)$, where $|\ell_k| < 1$, $\forall k \in \{2, \ldots, n\}$. Let $A = \mathrm{diag}\big(1, 1 + \sqrt{1-\ell_2^2}, \ldots, 1 + \sqrt{1-\ell_n^2}\big)$ and $B = A^{-1}$. Consider the sequences $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ generated by applying CPA to $0 \in Ax + L^\top BLx$ with $\gamma = \tau = 1$ and constant λ, where s^k is defined as in (28). Then, the following assertions hold.

(i) The associated primal-dual operator T_{PD} has a $\left(\frac{1}{2}\mathrm{I}_n \oplus \frac{1}{2}\mathrm{I}_n\right)$-oblique weak Minty solution at $(0, 0) = \mathrm{zer}\, T_{\mathrm{PD}}$.

(ii) By Theorem 3.1(iv) and Proposition 3.2, both $(\Pi_{\mathcal{R}(P)} z^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ converge to zero if λ is selected according to Relaxation parameter rule I, i.e., if

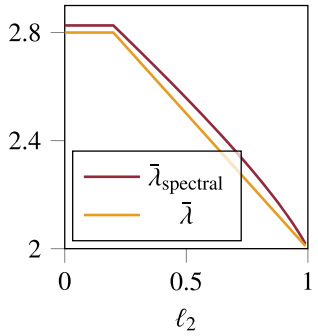

Fig. 4 The upper bounds $\bar{\lambda}$ and $\bar{\lambda}_{\text{spectral}}$ for Example 3.2, where $\bar{\lambda}_{\text{spectral}}$ is obtained by examining the spectral radius of the algorithmic operator $\Pi_{\mathcal{R}(P)}\left(\mathrm{I}+\lambda\big((P+T_{\text{PD}})^{-1}P-I\big)\right)$ (see (50))

$\lambda \in (0, \bar{\lambda})$, where

$$\begin{aligned}\bar{\lambda} &= 2\left(1+\tfrac{1}{2\gamma}\beta_{\text{P}}+\tfrac{1}{2\tau}\beta_{\text{D}}-\theta_{\gamma\tau}(\max\{|\ell_2|,\ldots,|\ell_n|\})\right)\\ &= 3-\max\{|\ell_2|,\ldots,|\ell_n|\}.\end{aligned}$$

(iii) Let $n = 3$, $\ell_2 \in (0, 1)$ and $\ell_3 = 1/5$. Then, the set of relaxation parameters for which the sequences $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ converge is almost entirely covered by Theorem 3.1(iv) (see Fig. 4).

4 Semimonotone Operators

In this section, we provide calculus rules for the class of (M, R)-semimonotone operators defined in Definition 1.2, generalizing the class of (μ, ρ)-semimonotone operators introduced in [20, Sec. 4]. Sufficient conditions for the convergence of CPA applied to (P-I) for (M, R)-semimonotone operators A and B will be provided in Sect. 5. The proofs of the calculus rules in this section are deferred to Appendix B.

For some choices of M and R, it follows from the Fenchel-Young inequality that all operators satisfy the definition of (M, R)-semimonotonicity, as stated below.

Proposition 4.1 *Let $M, R \in \mathbb{S}^n$. If $M \prec 0$, $R \prec 0$ and $M \preceq \frac{1}{4}R^{-1}$, then all operators $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ satisfy the definition of (M, R)-semimonotonicity.*

The following proposition demonstrates that the maximality of a semimonotone operator is preserved when the semimonotonicity moduli M and R are tightened.

Proposition 4.2 (preservation of maximality) *Let operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (M, R)-semimonotone. If T is maximally (M', R')-semimonotone for some $M' \preceq M$ and $R' \preceq R$, then T is maximally (M, R)-semimonotone.*

In what follows, various basic properties of (M, R)-semimonotone operators will be provided. For instance, by definition, their inverses belong to the same class of

operators, with the roles of M and R reversed. Additionally, the following proposition analyzes scaling and shifting of semimonotone operators, as well as the cartesian product of two semimonotone operators.

Proposition 4.3 (inverting, shifting, scaling and cartesian product) *Let operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (maximally) (M_A, R_A)-semimonotone [at $(x'_A, y'_A) \in \operatorname{gph} A$] and operator $B : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be (maximally) (M_B, R_B)-semimonotone [at $(x'_B, y'_B) \in \operatorname{gph} B$]. Let $\alpha \in \mathbb{R}_{++}$. Then, the following hold.*

(i) *The inverse operator A^{-1} is (maximally) (R_A, M_A)-semimonotone [at $(y'_A, x'_A) \in \operatorname{gph} A^{-1}$].*
(ii) *For all $u, w \in \mathbb{R}^n$, operator $T(x) := w + \alpha A(x + u)$ is (maximally) $(\alpha M_A, \alpha^{-1} R_A)$-semimonotone [at $(x'_A - u, w + \alpha y'_A)$].*
(iii) *Operator $T := A \times B$ is (maximally) $\big(M_A \oplus M_B, R_A \oplus R_B\big)$-semimonotone [at $(x', y') \in \operatorname{gph} T$ where $x' = (x'_A, x'_B)$ and $y' = (y'_A, y'_B)$].*

In Definition 1.2, there is some freedom in selecting the matrices M and R, which might lead to a tradeoff between both. One particular class of operators for which this is true is the class of linear operators. This is summarized in the following proposition, which generalizes [6, Prop. 5.1] for μ-monotone and ρ-comonotone operators and [20, Prop. 4.5] for (μ, ρ)-semimonotone operators.

Proposition 4.4 (linear operator) *Let $D \in \mathbb{R}^{n \times n}$ and let $M, R \in \mathbb{S}^n$. Then, D is (M, R)-semimonotone if and only if $\frac{1}{2}(D + D^\top) - M - D^\top R D \succeq 0$.*

Given a certain matrix D and a desired semimonotonicity modulus M, it might be difficult to determine whether there exists an R satisfying $D^\top R D \preceq \frac{1}{2}(D + D^\top) - M$, as this corresponds to solving a linear matrix inequality (LMI). The study of LMIs in general form has been extensively explored within the control and systems theory communities, leading to well-known results such as the Kalman–Yakubovich–Popov lemma, Finsler's lemma and the (nonstrict) projection lemma [4, 10, 26, 35]. In this work, we rely on a particular result for LMIs of the form $D^\top X D \preceq Y$, which is due to [55, 56] and relies upon the classical results from [3, 29, 42] for the linear matrix equality $D^\top X D = Y$.

Proposition 4.5 (symmetric solution of $D^\top X D \preceq Y$) *Let $D \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{S}^n$. Then, the following hold.*

(i) *The set of solutions $C := \{X \in \mathbb{S}^m \mid D^\top X D \preceq Y\}$ is nonempty if and only if*

$$\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)} \succeq 0 \quad \text{and} \quad \operatorname{rank}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)}) = \operatorname{rank}(\Pi_{\mathcal{N}(D)} Y). \tag{32}$$

(ii) *If (32) holds, then $X^\star \in C$, where*

$$X^\star = \begin{bmatrix} 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} -Y & D^\top \\ D & 0 \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix} = (D^\dagger)^\top \left(Y - Y \Pi_{\mathcal{N}(D)} (\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y \right) D^\dagger. \tag{33}$$

Moreover, $D^\top X D \preceq D^\top X^\star D \preceq Y$ *for all* $X \in C$.

(iii) *If the matrix equation* $D^\top X D = Y$ *is consistent, i.e. if* $\mathcal{R}(Y) \subseteq \mathcal{R}(D^\top)$, *then* $X^\star = (D^\dagger)^\top Y D^\dagger$ *is the solution of* $D^\top X D = Y$ *with minimal trace* $\operatorname{tr} X^2$.

Applying this result to Proposition 4.4, the following corollary for linear operators is obtained.

Corollary 4.1 *(*linear operator*) Let* $D \in \mathbb{R}^{n\times n}$ *and* $M \in \mathbb{S}^n$. *Then, the following hold.*

(i) *There exists* $R \in \mathbb{S}^n$ *such that* D *is* (M, R)*-semimonotone if and only if*

$$\Pi_{\mathcal{N}(D)} M \Pi_{\mathcal{N}(D)} \preceq 0 \text{ and } \operatorname{rank}(\Pi_{\mathcal{N}(D)} M \Pi_{\mathcal{N}(D)}) = \operatorname{rank}\left(\Pi_{\mathcal{N}(D)} (\tfrac{1}{2} D - M)\right). \tag{34}$$

(ii) *If (34) holds, then* D *is* $(M, R^\star)$*-semimonotone, where*

$$R^\star = \begin{bmatrix} 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} M - \frac{1}{2}(D + D^\top) & D^\top \\ D & 0 \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix}.$$

In particular, when D *is either symmetric or skew-symmetric, it holds that*

$$R^\star = 2(D + D^\top)^\dagger - {D^\dagger}^\top M D^\dagger + (D^\dagger)^\top M \Pi_{\mathcal{N}(D)} (\Pi_{\mathcal{N}(D)} M \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} M D^\dagger.$$

Note that $R^\star$ can be seen as an optimal choice for R, as it solves the LMI from Proposition 4.4 as tightly as possible. A second consequence of Proposition 4.5 is the following result on operators of the form $DTD^\top$.

Corollary 4.2 (semimonotonicity of $DTD^\top$) *Let* $D \in \mathbb{R}^{n\times m}$, $M, Y \in \mathbb{S}^m$ *and let operator* $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ *be* (M, Y)*-semimonotone [at* $(D^\top x', y') \in \operatorname{gph} T$*]. If (32) holds for* D *and* T, *then* $DTD^\top$ *is* $(DMD^\top, X^\star)$*-semimonotone [at* (x', Dy')*], where* $X^\star$ *is given by (33).*

Leveraging the previous result for the semimonotonicity of $DTD^\top$, the semimonotonicity of the sum and parallel sum of two semimonotone operators is investigated next. Recall that the parallel sum of two operators $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined as $A \,\square\, B := (A^{-1} + B^{-1})^{-1}$. We shall also use the parallel sum between symmetric matrices $X, Y \in \mathbb{S}^n$ as defined below.

Definition 4.1 (*parallel sum between symmetric matrices*) Let $X, Y \in \mathbb{S}^n$. We say that X and Y are parallel summable if $\mathcal{R}(X) \subseteq \mathcal{R}(X+Y)$ or equivalently $\mathcal{R}(Y) \subseteq \mathcal{R}(X+Y)$.

For parallel summable matrices X and Y, their parallel sum is defined as [37, Cor. 9.2.5]

$$X \square Y := X(X+Y)^{\dagger}Y = Y(X+Y)^{\dagger}X = X - X(X+Y)^{\dagger}X = Y - Y(X+Y)^{\dagger}Y. \tag{35}$$

For scalars $\alpha, \beta \in \mathbb{R}$, we say that α and β are parallel summable if either $\alpha = \beta = 0$ or $\alpha + \beta \neq 0$ and their parallel sum is defined as

$$\alpha \square \beta := \alpha(\alpha+\beta)^{\dagger}\beta = \begin{cases} 0, & \text{if } \alpha = \beta = 0, \\ \frac{\alpha\beta}{\alpha+\beta}, & \text{otherwise.} \end{cases}$$

In addition. consider the following set, which will be referred to as the effective domain of the parallel sum between two symmetric matrices.

Definition 4.2 (*effective domain of parallel sum*) The set

$$\mathrm{dom}_{\square} := \{(X, Y) \in \mathbb{S}^n \times \mathbb{S}^n \,|\, X + Y \succeq 0,\ X \text{ and } Y \text{ are parallel summable}\} \tag{36}$$

is the effective domain of the parallel sum between two symmetric (possibly indefinite) matrices. Let $X = \alpha \mathrm{I}_n$ and $Y = \beta \mathrm{I}_n$ where $\alpha, \beta \in \mathbb{R}$. Then, $(X, Y) \in \mathrm{dom}_{\square}$ reduces to $(\alpha, \beta) \in \mathrm{dom}_{\square} = \{(\alpha, \beta) | \alpha + \beta > 0 \text{ or } \alpha = \beta = 0\}$.

In the upcoming proposition, it is shown that the sum and parallel sum of two semimonotone operators are also semimonotone operators, if the involved semimonotonicity matrices belong to the effective domain of the parallel sum. This result generalizes [20, Prop. 4.7] for the sum of two (μ, ρ)-semimonotone operators.

Proposition 4.6 (sum and parallel sum) *Let operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (M_A, R_A)-semimonotone [at $(x'_A, y'_A) \in \mathrm{gph}\, A$] and operator $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (M_B, R_B)-semimonotone [at $(x'_B, y'_B) \in \mathrm{gph}\, B$].*

(i) *If $(R_A, R_B) \in \mathrm{dom}_{\square}$ [and $x'_A = x'_B =: x'$], then $A + B$ is $(M_A + M_B, R_A \square R_B)$-semimonotone [at $(x', y'_A + y'_B)$].*
(ii) *If $(M_A, M_B) \in \mathrm{dom}_{\square}$ [and $y'_A = y'_B =: y'$], then $A \square B$ is $(M_A \square M_B, R_A + R_B)$-semimonotone [at $(x'_A + x'_B, y')$].*

When one of the two involved operators is linear, more precise statements for the resulting semimonotonicity matrices can be derived. For instance, consider the following lemma for the sum of a semimonotone operator and a skew-symmetric matrix. This result will be used later in Theorem 5.1 for analyzing the primal-dual operator T_{PD}.

Lemma 4.1 (sum with skew-symmetric matrix) *Let $D \in \mathbb{R}^{n \times n}$ be a skew-symmetric matrix and operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be $(D^{\top} M D, R + R')$-semimonotone [at $(x', y') \in \mathrm{gph}\, T$], where $\mathcal{R}(R') \subseteq \mathcal{N}(D)$ and $(M, R) \in \mathrm{dom}_{\square}$. Then, $T + D$ is $(0, R' + M \square R)$-semimonotone [at $(x', y' + Dx')$].*

The calculus rules presented in this section can be used both to generate and to identify semimonotone operators. We remark that it is also possible to confirm semimonotonicity of operators either directly by definition or numerically through their *scaled relative graph*, as detailed in [45].

5 Chambolle–Pock for Semimonotone Operators

In Sect. 3, convergence of CPA was established under an oblique weak Minty assumption on the underlying primal-dual operator. This section provides a set of sufficient conditions for the convergence of CPA for composite inclusion problems involving semimonotone operators.

5.1 Existence of Oblique Weak Minty Solutions

The main tool for establishing simplified conditions for CPA for semimonotone operators is the following calculus rule, which connects the semimonotonicity of the individual operators A and B to the existence of V-oblique weak Minty solutions of the primal-dual operator T_{PD}.

Theorem 5.1 (primal-dual operator) *In the primal-dual inclusion (PD-I), suppose that there exists a nonempty set* $\mathcal{S}^\star \subseteq \operatorname{zer} T_{\mathrm{PD}}$ *and matrices* $R_A, R'_A, R_B \in \mathbb{S}^n$ *and* $M_A, M_B, M'_B \in \mathbb{S}^m$ *such that for every* $z^\star = (x^\star, y^\star) \in \mathcal{S}^\star$ *the following hold.*

(i) $\mathcal{R}(R_A) \subseteq \mathcal{R}(L^\top)$, $\mathcal{R}(R'_A) \subseteq \mathcal{N}(L)$, $\mathcal{R}(M_B) \subseteq \mathcal{R}(L)$ *and* $\mathcal{R}(M'_B) \subseteq \mathcal{N}(L^\top)$.
(ii) $(M_A, M_B) \in \operatorname{dom}_{\square}$ *and* $(R_A, R_B) \in \operatorname{dom}_{\square}$.
(iii) *Operator* A *is* $(L^\top M_A L, R_A + R'_A)$*-semimonotone at* $(x^\star, -L^\top y^\star) \in \operatorname{gph} A$.
(iv) *Operator* B *is* $(M_B + M'_B, L R_B L^\top)$*-semimonotone at* $(Lx^\star, y^\star) \in \operatorname{gph} B$.

Then, the primal-dual operator T_{PD} *has* $((R_A \square R_B + R'_A) \oplus (M_A \square M_B + M'_B))$*-oblique weak Minty solutions at* $\mathcal{S}^\star$.

Proof Let $(x^\star, y^\star) \in \mathcal{S}^\star$ and decompose the primal-dual operator as $T_{\mathrm{PD}} = T + D$, where $T := A \times B^{-1}$ and $D(x, y) := (L^\top y, -Lx)$. By Proposition 4.3(i), the inverse operator B^{-1} is $(L R_B L^\top, M_B + M'_B)$-semimonotone at $(y^\star, Lx^\star) \in \operatorname{gph} B^{-1}$. Therefore, it follows from Proposition 4.3(iii) for the Cartesian product of two semimonotone operators that T is $\big(D^\top (M_A \oplus R_B) D, (R_A + R'_A) \oplus (M_B + M'_B)\big)$-semimonotone at $\big((x^\star, y^\star), (-L^\top y^\star, Lx^\star)\big) \in \operatorname{gph} T$. Having established the semimonotonicity of T, it follows from Lemma 4.1 and skew-symmetry of D that $T_{\mathrm{PD}} = T + D$ is $\big(0, (R_A \square R_B + R'_A) \oplus (M_A \square M_B + M'_B)\big)$-semimonotone at $((x^\star, y^\star), 0)$. The claim then follows by Definition 1.1. □

Suppose that the underlying assumptions from Theorem 5.1 hold, in which case $\mathcal{R}(R_A \square R_B) \subseteq \mathcal{R}(R_A) \subseteq \mathcal{R}(L^\top)$ and $\mathcal{R}(M_A \square M_B) \subseteq \mathcal{R}(M_B) \subseteq \mathcal{R}(L)$. Then, by virtue of (7) and the particular form of $V = \big((R_A \square R_B + R'_A) \oplus (M_A \square M_B + M'_B)\big)$ from Theorem 5.1, the primal-dual operator T_{PD} has V-oblique weak Minty solutions

at $\mathcal{S}^\star$, with V given by (8) and

$$\begin{aligned} \beta_{\mathrm{P}} &= \lambda_{\min}(X^\top (R_A \square R_B) X), \quad & \beta'_{\mathrm{P}} &= \lambda_{\min}(X'^\top R'_A X'), \\ \beta_{\mathrm{D}} &= \lambda_{\min}(Y^\top (M_A \square M_B) Y), \quad & \beta'_{\mathrm{D}} &= \lambda_{\min}(Y'^\top M'_B Y'), \end{aligned} \tag{37}$$

where X, X', Y, Y' are defined as in (6). Hence, Assumption II.A3 holds if the parameters from (37) satisfy condition (10).

As an implication of Theorem 5.1, one can easily verify the underlying assumptions of Theorem 3.1. For instance, the following proposition applies to constrained, nonconvex quadratic optimization problems, with its proof provided in Appendix B.

Proposition 5.1 (constrained QP) *Consider the quadratic program*

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \ \tfrac{1}{2} x^\top Q x + q^\top x \quad \text{subject to} \quad Lx \in C, \tag{38}$$

where $Q \in \mathbb{S}^n$, $q \in \mathbb{R}^n$. $L \in \mathbb{R}^{m\times n}$ *and* $C := \{x \in \mathbb{R}^m \mid l_i \le x_i \le u_i,\ i = 1, \ldots, m\}$, *where* $l, u \in \mathbb{R}^m$. *The associated first-order optimality condition is given by* $0 \in Ax + L^\top B L x$, *where* $A : x \mapsto Qx + q$ *and* $B := N_C$. *Suppose that* L *is either full column rank, or that* $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$ *and* $\Pi_{\mathcal{N}(L)} Q \Pi_{\mathcal{N}(L)} \succeq 0$. *Then, the following hold.*

(i) *Operator* A *is* $(L^\top M_A L, R'_A)$*-semimonotone, where* $M_A = L^{\dagger\top} Q L^\dagger$ *and* $R'_A = \Pi_{\mathcal{N}(L)} Q^\dagger \Pi_{\mathcal{N}(L)} \succeq 0$.

(ii) *Let* $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ *and* $M_B(y^\star) := \operatorname{diag}\left(\frac{|y_1^\star|}{u_1 - l_1}, \ldots, \frac{|y_n^\star|}{u_n - l_n}\right)$. *Then, operator* B *is* $\left(M_B(y^\star), 0\right)$*-semimonotone at* $(Lx^\star, y^\star) \in \operatorname{gph} B$.

Moreover, if there exists a primal-dual pair $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ *satisfying* $(M_A, M_B(y^\star)) \in \operatorname{dom}_\square$, *the following hold.*

(iii) *Let* X' *and* Y *be as in (6). The pair* $(x^\star, y^\star)$ *is a* V*-oblique weak Minty solution of* T_{PD} *with* V *given by (8), where* $\beta_{\mathrm{P}} = 0$, $\beta_{\mathrm{D}} = \lambda_{\min}\left(Y^\top (M_A \square M_B(y^\star))\, Y\right)$,

$$\beta'_{\mathrm{P}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = n, \\ \lambda_{\min}\left(X'^\top Q^\dagger X'\right) \ge 0, & \text{otherwise,} \end{cases}$$

and

$$\beta'_{\mathrm{D}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = m, \\ 0, & \text{otherwise.} \end{cases}$$

(iv) *Let the stepsizes* γ, τ *and fixed relaxation parameter* λ *satisfy*

$$\gamma \in \left(0, \tfrac{1}{[\beta_{\mathrm{D}}]_- \|L\|^2}\right),\ \tau \in \left([\beta_{\mathrm{D}}]_-, \tfrac{1}{\gamma\|L\|^2}\right),\ \lambda \in \left(0, 2(1 + \min\{-\tfrac{1}{\tau}[\beta_{\mathrm{D}}]_-, \tfrac{2}{\gamma}\beta'_{\mathrm{P}}\})\right). \tag{39}$$

Then, Assumption II, Stepsize rule I, and Relaxation parameter rule I hold, the resolvents $J_{\gamma A}$ and $J_{\tau B^{-1}}$ are (single-valued) continuous and the convergence properties of CPA follow from Theorem 3.1.

Notably, Proposition 5.1 covers examples of nonconvex quadratic programs for which the matrix L is rank-deficient. For instance, consider the following two examples with an indefinite Q matrix, where L is full row rank in the first example, while it is neither full row nor full column rank in the second. The proofs are deferred to Appendix B.

Example 5.1 Consider the QP from (38), where $Q = \operatorname{diag}(1, -1, 2)$, $q = [\,-1\ 1\ -1\,]^\top$, $L = \left[\begin{smallmatrix} 1 & 1/4 & 0 \\ 0 & 1 & 0 \end{smallmatrix}\right]$ and $C := \{x \in \mathbb{R}^2 \mid 2 \leq x_i \leq 4, i = 1, 2\}$. Let operators A and B be defined as in Proposition 5.1. Then, the global minimizer of (38) is given by $x^\star = [\,1\ 4\ 1/2\,]^\top$ and the following hold.

(i) A is $\big(\operatorname{diag}(1, -1, 0), \operatorname{diag}(0, 0, 1/2)\big)$-semimonotone.
(ii) B is $\big(\operatorname{diag}(0, 3/2), 0\big)$-semimonotone at $(Lx^\star, -L^{\dagger\top} Ax^\star) = \left(\left[\begin{smallmatrix} 2 \\ 4 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 3 \end{smallmatrix}\right]\right) \in \operatorname{gph} B$.
(iii) The pair $(x^\star, -L^{\dagger\top} Ax^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ is a V-oblique weak Minty solution of T_{PD} with V given by (8), where $\beta_{\mathrm{P}} = 0$, $\beta_{\mathrm{D}} = -3$, $\beta'_{\mathrm{P}} = 1/2$ and $\beta'_{\mathrm{D}} = \infty$.
(iv) The sequence $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ generated by CPA with fixed relaxation parameter λ converges for $\gamma \in (0, 0.26)$, $\tau \in (3, \frac{0.779}{\gamma})$ and $\lambda \in (0, 2 - \frac{6}{\tau})$.

Example 5.2 Consider the QP from (38), where $Q = \operatorname{diag}(-3, -2, 1)$, $q = [\,0\ 1\ 0\,]^\top$, $L = \left[\begin{smallmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{smallmatrix}\right]$ and $C = \{x \in \mathbb{R}^2 \mid 1/2 \leq x_i \leq 1,\ i = 1, 2, 3\}$. Let operators A and B be defined as in Proposition 5.1. Then, the global minimizer is given by $x^\star = [\,1\ 0\ 0\,]^\top$ and the following hold.

(i) A is $\big(-L^\top L, \operatorname{diag}(0, 0, 1)\big)$-semimonotone.
(ii) B is $\big(2, 0\big)$-semimonotone at $(Lx^\star, -L^{\dagger\top} Ax^\star) = \big([\,1\ 1\ 1\,]^\top, [\,1\ 1\ 1\,]^\top\big) \in \operatorname{gph} B$.
(iii) The pair $(x^\star, -L^{\dagger\top} Ax^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ is a V-oblique weak Minty solution of T_{PD} with V given by (8), where $\beta_{\mathrm{P}} = 0$, $\beta_{\mathrm{D}} = -2$, $\beta'_{\mathrm{P}} = 1$ and $\beta'_{\mathrm{D}} = 0$.
(iv) The sequence $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$ generated by CPA with fixed relaxation parameter λ converges for $\gamma \in (0, \frac{1}{6})$, $\tau \in (2, \frac{1}{3\gamma})$ and $\lambda \in (0, 2 - \frac{4}{\tau})$.

5.2 Sufficient Conditions for Convergence of CPA

Theorem 5.1(i) imposes range conditions on the semimonotonicity matrices of A and B. In this subsection, it is shown that this can be achieved by imposing a certain structure on these matrices. In particular, consider the following set of assumptions.

Assumption III In problem (P-I), suppose that $\operatorname{zer} T_{\mathrm{PD}}$ is nonempty and that there exists a nonempty set $\mathcal{S}^\star \subseteq \operatorname{zer} T_{\mathrm{PD}}$ such that for every $z^\star = (x^\star, y^\star) \in \mathcal{S}^\star$ it holds that A is $(\mu_A L^\top L, \rho_A \mathrm{I}_n)$-semimonotone at $(x^\star, -L^\top y^\star) \in \operatorname{gph} A$, B is $(\mu_B \mathrm{I}_m, \rho_B L L^\top)$-semimonotone at $(Lx^\star, y^\star) \in \operatorname{gph} B$ and the semimonotonicity moduli $(\mu_A, \mu_B, \rho_A, \rho_B) \in \mathbb{R}^4$ satisfy either one of the following conditions.

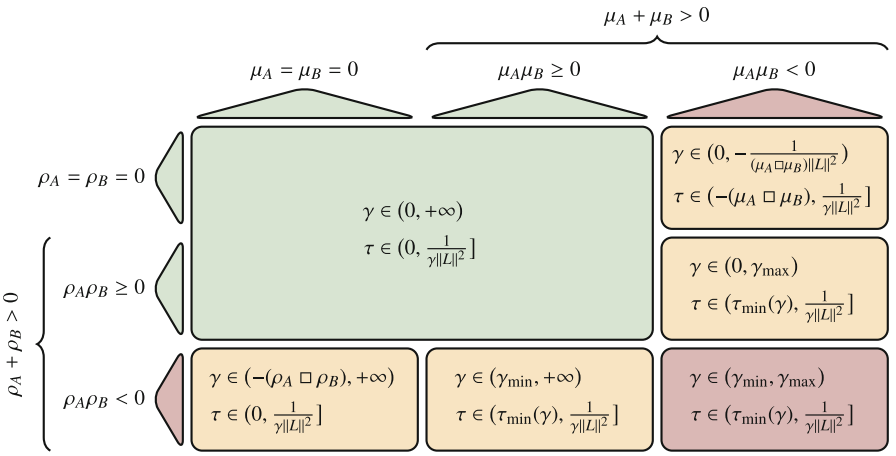

Fig. 5 Range of the stepsizes γ and τ for CPA involving semimonotone operators

(i) (either) $\mu_A = \mu_B = 0$ and $\rho_A = \rho_B = 0$.
(ii) (or) $\mu_A + \mu_B > 0$ and $\rho_A = \rho_B = 0$.
(iii) (or) $\rho_A + \rho_B > 0$ and $\mu_A = \mu_B = 0$.
(iv) (or) $\mu_A + \mu_B > 0$, $\rho_A + \rho_B > 0$ and $[\mu_A \square \mu_B]_-[\rho_A \square \rho_B]_- < \frac{1}{4\|L\|^2}$.

Assumption III provides a set of sufficient conditions on the operators A and B so that the primal-dual operator admits oblique weak Minty solutions. This key result is stated in the following corollary.

Corollary 5.1 *Suppose that Assumption III holds. Then, the primal-dual operator T_{PD} has V-oblique weak Minty solutions at $\mathcal{S}^\star$, where V is given by (8) and*

$$\beta_{\mathrm{P}} = \rho_A \square \rho_B, \quad \beta_{\mathrm{D}} = \mu_A \square \mu_B,$$
$$\beta'_{\mathrm{P}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = n, \\ \rho_A, & \text{if } \operatorname{rank} L < n, \end{cases}, \beta'_{\mathrm{D}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = m, \\ \mu_B, & \text{if } \operatorname{rank} L < m. \end{cases}$$

Proof Note that $(\mu_A, \mu_B) \in \mathrm{dom}\,\square$ and $(\rho_A, \rho_B) \in \mathrm{dom}\,\square$ by Assumption III and that $\rho_A \mathrm{I}_n = \rho_A \Pi_{\mathcal{R}(L^\top)} + \beta'_{\mathrm{P}} \Pi_{\mathcal{N}(L)}$ and $\mu_B \mathrm{I}_m = \mu_B \Pi_{\mathcal{R}(L)} + \beta'_{\mathrm{D}} \Pi_{\mathcal{N}(L^\top)}$. Therefore, the claim follows directly from Theorem 5.1, with $M_A = \mu_A \Pi_{\mathcal{R}(L)}$, $R_A = \rho_A \Pi_{\mathcal{R}(L^\top)}$, $R'_A = \beta'_{\mathrm{P}} \Pi_{\mathcal{N}(L)}$, $M_B = \mu_B \Pi_{\mathcal{R}(L)}$, $M'_B = \beta'_{\mathrm{D}} \Pi_{\mathcal{N}(L^\top)}$ and $R_B = \rho_B \Pi_{\mathcal{R}(L^\top)}$. □

Leveraging this key result, the convergence of CPA for semimonotone operators is analyzed in Corollary 5.2, for which the stepsize rule corresponds to a simple look-up table. The proof is deferred to Appendix B.

Stepsize rule II *The stepsizes γ and τ satisfy the bounds provided in Fig. 5, where $\gamma_{\min}$, $\gamma_{\max}$ and $\tau_{\min}(\gamma)$ are defined as in (11) with $\beta_{\mathrm{D}} = \mu_A \square \mu_B$ and $\beta_{\mathrm{P}} = \rho_A \square \rho_B$.*

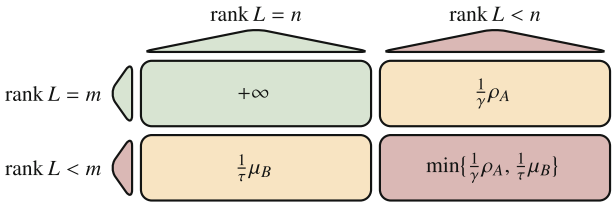

Fig. 6 Definition of η' in Relaxation parameter rule II

Relaxation parameter rule II *Let η' be defined as in Fig. 6 and define*

$$\Delta_{\gamma,\tau} := \tfrac{1}{2\gamma}(\rho_A \,\square\, \rho_B) + \tfrac{1}{2\tau}(\mu_A \,\square\, \mu_B) \quad \text{and}$$
$$\theta_{\gamma\tau}(\sigma) := \sqrt{\Delta_{\gamma,\tau}^2 + (\mu_A \,\square\, \mu_B)(\rho_A \,\square\, \rho_B)\sigma^2}.$$

The sequence $(\lambda_k)_{k\in\mathbb{N}}$ satisfies $\lambda_k \in \big(0, 2(1+\bar{\eta})\big)$ and $\liminf_{k\to\infty} \lambda_k\big(2(1+\bar{\eta})-\lambda_k\big) > 0$, where

$$\bar{\eta} := \begin{cases} \left\{\begin{array}{ll} \Delta_{\gamma,\tau} - \theta_{\gamma\tau}(\|L\|), & \text{if } \mu_A\mu_B\rho_A\rho_B \geq 0 \\ \min\left\{\Delta_{\gamma,\tau} - \theta_{\gamma\tau}(\sigma_d), \eta'\right\}, & \text{otherwise} \end{array}\right\} & \text{if } \gamma\tau < \frac{1}{\|L\|^2}, \\ \left\{\begin{array}{ll} 2\Delta_{\gamma,\tau}, & \text{if } \max\{\mu_A\mu_B, \rho_A\rho_B\} \leq 0 \\ \min\left\{2\Delta_{\gamma,\tau}, \eta'\right\}, & \text{otherwise} \end{array}\right\} & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d = 1, \\ \left\{\begin{array}{ll} 2\Delta_{\gamma,\tau}, & \text{if } \max\{\mu_A\mu_B, \rho_A\rho_B\} \leq 0 \\ \Delta_{\gamma,\tau} - \theta_{\gamma\tau}(\sigma_2), & \text{if } \min\{\mu_A\mu_B, \rho_A\rho_B\} > 0 \\ \min\left\{\Delta_{\gamma,\tau} - \theta_{\gamma\tau}(\sigma_d), \eta'\right\}, & \text{otherwise} \end{array}\right\} & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d > 1. \end{cases}$$

Corollary 5.2 (convergence of CPA under semimonotonicity) *Suppose that Assumption II.A1, Assumption II.A2 and Assumption III hold, that γ and τ are selected according to Stepsize rule II and that the relaxation sequence $(\lambda_k)_{k\in\mathbb{N}}$ is selected according to Relaxation parameter rule II. Then, all the claims of Theorem 3.1 hold.*

Corollary 5.2 serves as a universal framework for analyzing the convergence of CPA, both in monotone and nonmonotone settings. Notably, it encompasses and extends many of the existing results in literature. Several examples are provided below.

Remark 5.1 (connection to existing theory) Case (i) of Assumption III can be interpreted as a relaxation of the classical monotonicity assumption for CPA [12], as it does not impose monotonicity between any two pairs in the graph of A and B. In case (ii) of Assumption III, a monotone problem is split in a nonmonotone fashion. In the optimization setting, this setting was studied in [38]. To see this, let g be a proper lsc μ_g-convex function with $\mu_g > 0$ and h be a proper lsc μ_h-convex function. Then, $A = \partial g$ is $(\mu_g/\|L\|^2 L^\top L, 0)$-semimonotone and $B = \partial h$ is $(\mu_h, 0)$-semimonotone. Then, Corollary 5.2 requires that $\mu_g/\|L\|^2 + \mu_h > 0$, which matches [38, Thm. 2.8]. Note that case (iii) of Assumption III can be interpreted as the dual counterpart of case (ii), as the assumptions of the latter hold for A and B in the primal inclusion problem

if and only if the assumptions of the former hold for B^{-1} and A^{-1} in the dual one (see (D-I)). Up to the knowledge of the authors, no particular instances of case (iv) of Assumption III have been covered in literature, even in the minimization setting.

In what follows, Corollary 5.2 will be applied to several examples previously discussed in this paper. First, note that the convergence result from Example 5.2(iv) can be more conveniently derived as a consequence of Corollary 5.2, since Assumption III is satisfied for this problem with $\mu_A = -1$, $\mu_B = 2$, $\rho_A = 1$ and $\rho_B = 0$. Next, we revisit the two theoretical examples of Sect. 3, this time under the lens of semimonotonicity. First of all, consider the linear inclusion problem from Example 3.1, where the parameters a, b, c, l are selected as in Example 3.1(iv).

Example 5.3 (saddle point problem (revisited)) Consider inclusion problem (29) with $a = 10$, $b = -\frac{1}{4}$, $c = 0$ and $\ell = 2$. Using Proposition 4.4, it follows that A is $\left(L^\top L, -\frac{1}{25}\mathrm{I}_n\right)$-semimonotone and B is $\left(-\frac{3}{10}\mathrm{I}_m, \frac{1}{5}L^\top L\right)$-semimonotone. By Corollary 5.2, the sequence $(z^k)_{k\in\mathbb{N}}$ generated by applying CPA to (29) with $\tau = \frac{1}{\gamma\|L\|^2}$ and fixed relaxation parameter λ converges for $\gamma \in (0.055, 0.528)$ and $\lambda \in \left(0, 2 - \frac{1}{10\gamma} - \frac{24\gamma}{7}\right)$.

The obtained range of stepsize parameters is only a subset of the tight range obtained in Example 3.1(iv). However, this should not come as a surprise, since part of the information about operators A and B is lost by analyzing them under the lens of semimonotonicity. This is also observed in the second example.

Example 5.4 (influence of singular values (revisited)) Consider the composite inclusion problem $0 \in Ax + L^\top BLx$ from Example 3.2. It follows from Proposition 4.4 that A is $\left(\frac{1}{2}L^\top L, \frac{1}{2}\mathrm{I}_n\right)$-semimonotone and B is $\left(\frac{1}{2}\mathrm{I}_n, \frac{1}{2}LL^\top\right)$-semimonotone. By Corollary 5.2 and Proposition 3.2, the sequences $(\Pi_{\mathcal{R}(P)}z^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ generated by applying CPA with $\gamma = \tau = 1$ and fixed relaxation parameter λ, converge to zero if λ is selected according to Relaxation parameter rule II, which reduces to $\lambda \in \left(0, 5/2 - 1/2\max\{|\ell_2|, \ldots, |\ell_n|\}\right)$.

In Example3.2(i), it was shown that T_{PD} has $(\frac{1}{2}\mathrm{I}_n \oplus \frac{1}{2}\mathrm{I}_n)$-oblique weak Minty solutions at $\operatorname{zer} T_{\mathrm{PD}}$. On the other hand, here it is shown that A is $(\frac{1}{2}L^\top L, \frac{1}{2}\mathrm{I}_n)$-semimonotone and B is $(\frac{1}{2}\mathrm{I}_n, \frac{1}{2}LL^\top)$-semimonotone. Applying Corollary 5.1, this implies that T_{PD} only has a $(\frac{1}{4}\mathrm{I}_n \oplus \frac{1}{4}\mathrm{I}_n)$-oblique weak Minty solution at $(0_n, 0_n) = \operatorname{zer} T_{\mathrm{PD}}$. By analyzing A and B under the lens of semimonotonicity, some additional looseness is inevitably introduced. More specifically, the information that A and B are linear and symmetric, that $A = B^{-1}$ and that $A = 1/2L^\top L + 1/2A^\top A$ are lost in this process.

Finally, for a more practical application of our theory, we refer the interested reader to [45, Prop. V.2], where Corollary 5.2 is applied to a nonmonotone common-emitter amplifier circuit.

6 Conclusion

In this work, convergence of the Chambolle–Pock algorithm (CPA) was established for a class of nonmonotone problems, characterized by an oblique weak Minty assumption on the associated primal-dual operator. To facilitate the verification of this underlying assumption, a generalization of the class of semimonotone operators (see [20]) was introduced, and sufficient conditions for the convergence of CPA were provided for inclusion problems involving operators belonging to this class. When restricting to minimization problems, our results reveal that for certain problem classes no explicit rank or condition number restriction on the linear mapping is required.

It would be interesting to explore if stronger results can be obtained when the operators are known to be subdifferentials. Other future research directions include extensions to the setting where the preconditioning is indefinite, allowing to cover the extended Chambolle–Pock stepsize range $\gamma\tau\|L\|^2 \leq 4/3$ from [5, 57], as well as analyzing other splitting methods in nonmonotone settings.

A Auxiliary Results

Fact 1 (solution of quadratic inequality) *Let $\beta_{\mathrm{P}}, \beta_{\mathrm{D}} \in \mathbb{R}$, $\|L\| > 0$, $\sigma_d \in (0, \|L\|]$, let δ be defined as in (9) and let $\gamma_{\min}$ and $\gamma_{\max}$ be defined as in Stepsize rule I. Then, the following hold.*

(i) *There exists a $\gamma > 0$ satisfying $\beta_{\mathrm{D}}\|L\|^2\gamma^2 + \delta\gamma + \beta_{\mathrm{P}} > 0$ iff $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$.*

(ii) *If $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$, then $\gamma > 0$ satisfies $\beta_{\mathrm{D}}\|L\|^2\gamma^2 + \delta\gamma + \beta_{\mathrm{P}} > 0$ if and only if γ lies within the (nonempty) stepsize interval provided in Fig. 1.*

Lemma A.1 *Let $\beta_{\mathrm{P}}, \beta_{\mathrm{D}} \in \mathbb{R}$, $\|L\| > 0$, $\sigma_d \in (0, \|L\|]$ and for any $\sigma \in [\sigma_d, \|L\|]$ define the set*

$$\Gamma(\sigma) := \left\{ (\gamma, \tau) \in \mathbb{R}^2_{++} \;\middle|\; \gamma\tau \in \left(0, 1/\|L\|^2\right] \text{ and } 1 + \tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma) > 0 \right\},$$

where $\theta_{\gamma\tau}(\cdot)$ is defined as in (12). Then, the following hold.

(i) *If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0$, then, for any $\sigma \in [\sigma_d, \|L\|]$, stepsizes (γ, τ) belonging to the nonempty set $\Gamma(\sigma)$ are as in Fig. 1.*

(ii) *If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} < 0$ and $\max\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0$, then, stepsizes (γ, τ) belonging to the nonempty set $\Gamma(\sigma_d)$ are as in Fig. 1.*

(iii) *If $\max\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} < 0$, then the set $\Gamma(\|L\|)$ is nonempty if and only if $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$, in which case stepsizes (γ, τ) belonging to the nonempty set $\Gamma(\|L\|)$ are as in Fig. 1.*

Proof Throughout the proof, we will use $\gamma_{\min}$, $\gamma_{\max}$ and $\tau_{\min}(\gamma)$ as defined as in (11). Let $\sigma \in [\sigma_d, \|L\|]$. Solving the square root inequality $1 + \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} > \theta_{\gamma\tau}(\sigma)$, it

follows that

$$\Gamma(\sigma) = \left\{ (\gamma,\tau) \in \mathbb{R}^2_{++} \;\middle|\; \underbrace{\gamma\tau \in \left(0, 1/\|L\|^2\right], \overbrace{1 + \tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}}}^{\Gamma_1 :=} > 0,}_{\Gamma_2(\sigma) :=} \; 1 + \tfrac{1}{\gamma}\beta_{\mathrm{P}} + \tfrac{1}{\tau}\beta_{\mathrm{D}} + \tfrac{1}{\gamma\tau}\beta_{\mathrm{P}}\beta_{\mathrm{D}}(1-\gamma\tau\sigma^2) > 0 \right\}.$$

Define $c_1(\sigma,\gamma) := \gamma(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma^2) + \beta_{\mathrm{P}}$ and $c_2(\gamma) := \beta_{\mathrm{D}}(\gamma + \beta_{\mathrm{P}})$, so that $\gamma\tau\Gamma_2(\sigma) = c_1(\sigma,\gamma)\tau + c_2(\gamma)$.

♠ A.1(i): If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0$, then $\Gamma_1 > 0$ and $\Gamma_2(\sigma) > 0$ since $1 - \gamma\tau\sigma^2 \geq 0$. Therefore, it follows by algebraic manipulation that $\Gamma(\sigma)$ is nonempty and equal to

$$\Gamma(\sigma) = \{(\gamma,\tau) \in \mathbb{R}^2 | \gamma \in (0, +\infty) \text{ and } \tau \in \left(0, 1/\gamma\|L\|^2\right]\}.$$

♠ A.1(ii): If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} < 0$ and $\max\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0$, then either $\frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}} \geq 0$, in which case by definition $\Gamma_1 > 0$, or $\frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}} < 0$, in which case $\Gamma_1 > \Gamma_2(\sigma_d)$ since $1 - \gamma\tau\sigma_d^2 \geq 0$. Therefore, regardless of the sign of $\frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}$ it holds that $\Gamma(\sigma_d)$ reduces to $\{(\gamma,\tau) \in \mathbb{R}^2_{++} | \gamma\tau \in \left(0, 1/\|L\|^2\right]$ and $\Gamma_2(\sigma_d) > 0\}$.

◇ $\beta_{\mathrm{P}} \geq 0$, $\beta_{\mathrm{D}} < 0$: Then, by construction $c_1(\sigma_d,\gamma) > 0$ and $c_2(\gamma) < 0$ for all $\gamma \in \mathbb{R}_+$. Consequently, $\Gamma_2(\sigma_d) > 0 \iff c_1(\sigma_d,\gamma)\tau + c_2(\gamma) > 0 \iff \tau > {}^{-c_2(\gamma)}/_{c_1(\sigma_d,\gamma)} = \tau_{\min}(\gamma)$ and the set $\Gamma(\sigma_d)$ is nonempty if for some $\gamma > 0$ it holds that

$$\begin{aligned} &\tau_{\min}(\gamma) < 1/\gamma\|L\|^2 \iff \\ &\qquad c_2(\gamma)\gamma\|L\|^2 + c_1(\sigma_d,\gamma) = \beta_{\mathrm{D}}\|L\|^2\gamma^2 + \left(1 + \beta_{\mathrm{P}}\beta_{\mathrm{D}}\left(\|L\|^2 - \sigma_d^2\right)\right)\gamma + \beta_{\mathrm{P}} > 0, \end{aligned} \tag{40}$$

which is guaranteed by Fact 1(i). Therefore, it follows from Fact 1(ii) that

$$\Gamma(\sigma_d) = \left\{ (\gamma,\tau) \in \mathbb{R}^2 \;\middle|\; \gamma \in \left(0, \gamma_{\max}\right) \text{ and } \tau \in \left(\tau_{\min}(\gamma), 1/\gamma\|L\|^2\right] \right\}.$$

◇ $\beta_{\mathrm{P}} < 0$, $\beta_{\mathrm{D}} \geq 0$: Observe that for all $\gamma \leq {}^{-\beta_{\mathrm{P}}}/_{(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2)} < -\beta_{\mathrm{P}}$ it holds that $c_1(\sigma_d,\gamma) \leq 0$ and $c_2(\gamma) < 0$. Therefore, if $\Gamma_2(\sigma_d) > 0$ then necessarily $\gamma > {}^{-\beta_{\mathrm{P}}}/_{(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2)}$, implying $c_1(\sigma_d,\gamma) > 0$, so that $\Gamma(\sigma_d)$ reduces to $\left\{ (\gamma,\tau) \in \mathbb{R}^2_{++} \;\middle|\; \gamma\tau \in \left(0, 1/\|L\|^2\right], \gamma > {}^{-\beta_{\mathrm{P}}}/_{(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2)} \text{ and } \tau > \tau_{\min}(\gamma) \right\}$. Ensuring that $\tau_{\min}(\gamma) < 1/\gamma\|L\|^2$, i.e., solving (40) as before, yields $\gamma \in (\gamma_{\min}, +\infty)$. Finally, by observing that $\gamma_{\min} > -\beta_{\mathrm{P}} > {}^{-\beta_{\mathrm{P}}}/_{(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2)}$, it follows that

$$\Gamma(\sigma_d) = \left\{ (\gamma,\tau) \in \mathbb{R}^2 \;\middle|\; \gamma \in \left(\gamma_{\min}, +\infty\right) \text{ and } \tau \in \left(\tau_{\min}(\gamma), 1/\gamma\|L\|^2\right] \right\}.$$

♠ A.1(iii): If $\max\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} < 0$, then $\Gamma_1 > 0$ for $(\gamma, \tau) \in \mathbb{R}^2_{++}$ if and only if $\gamma > -\beta_{\mathrm{P}}/2$ and $\tau > -\gamma\beta_{\mathrm{D}}/(2\gamma+\beta_{\mathrm{P}})$. As a result, the set $\Gamma(\|L\|)$ is empty when $\beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 1/\|L\|^2$, as in this case the inequality $-\gamma\beta_{\mathrm{D}}/(2\gamma+\beta_{\mathrm{P}}) < 1/\gamma\|L\|^2$ does not have a positive solution for $\gamma > -\beta_{\mathrm{P}}/2$. Consider the following cases, assuming that $\beta_{\mathrm{P}}\beta_{\mathrm{D}} < 1/\|L\|^2$.

- ◇ $\gamma \in (-\beta_{\mathrm{P}}/2, -\beta_{\mathrm{P}}/(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2))$: Then, $c_1(\|L\|, \gamma) < 0$ and thus $\Gamma_2(\|L\|) > 0$ if and only if $\tau < \tau_{\min}(\gamma)$. Since it is easy to verify that in this case $-\gamma\beta_{\mathrm{D}}/(2\gamma+\beta_{\mathrm{P}}) > \tau_{\min}(\gamma)$, no such γ belong to the set $\Gamma(\|L\|)$.
- ◇ $\gamma = -\beta_{\mathrm{P}}/(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2)$: Then, $c_1(\|L\|, \gamma) = 0$ and $c_2(\gamma) < 0$, so that $\Gamma_2(\|L\|) < 0$ for all $\tau \in \mathbb{R}$.
- ◇ $\gamma > -\beta_{\mathrm{P}}/(1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2)$: Then, $c_1(\|L\|, \gamma) > 0$ and thus $\Gamma_2(\|L\|) > 0$ if and only if $\tau > \tau_{\min}(\gamma)$, which is by construction larger than $-\gamma\beta_{\mathrm{D}}/(2\gamma+\beta_{\mathrm{P}})$. The set $\Gamma(\sigma_d)$ is nonempty if for some $\gamma > 0$ it holds that $\tau_{\min}(\gamma) < 1/\gamma\|L\|^2$, which holds by Fact 1(i) if and only if $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < 1/4\|L\|^2$. Finally, noting that $\gamma_{\min} > -\beta_{\mathrm{P}}/2$, it follows from Fact 1(ii) that

$$\Gamma(\|L\|) = \left\{(\gamma, \tau) \in \mathbb{R}^2 \;\middle|\; \gamma \in \left(\gamma_{\min}, \gamma_{\max}\right) \text{ and } \tau \in \left(\tau_{\min}(\gamma), 1/\gamma\|L\|^2\right]\right\}.$$

Since these sets corresponds to the stepsize ranges provided by Fig. 1, the proof is completed.

Lemma A.2 *Let* $D := \left[\begin{smallmatrix} \mathrm{I}_n \\ \mathrm{I}_n \end{smallmatrix}\right]$, $(Y_1, Y_2) \in \mathrm{dom}\,\square$ *and define* $Y = Y_1 \oplus Y_2$. *Then, (32) holds and* $X^\star$ *as defined in (33) is equal to* $Y_1 \,\square\, Y_2$.

Proof Let $E = \frac{1}{2}\left[\begin{smallmatrix} 1 \\ -1 \end{smallmatrix}\right]$. Observe that $D^\dagger = \frac{1}{2}\left[\begin{smallmatrix} \mathrm{I}_n \\ \mathrm{I}_n \end{smallmatrix}\right]$ and $\Pi_{\mathcal{N}(D)} = \frac{1}{2}\left[\begin{smallmatrix} \mathrm{I}_n & -\mathrm{I}_n \\ -\mathrm{I}_n & \mathrm{I}_n \end{smallmatrix}\right]$, so that $\Pi_{\mathcal{N}(D)}Y = (E \otimes \mathrm{I}_n)\left[\begin{smallmatrix} Y_1 & -Y_2 \end{smallmatrix}\right]$ and $\Pi_{\mathcal{N}(D)}Y\Pi_{\mathcal{N}(D)} = (EE^\top) \otimes (Y_1 + Y_2)$. Therefore, $\mathrm{rank}(\Pi_{\mathcal{N}(D)}Y) = \mathrm{rank}\left[\begin{smallmatrix} Y_1 & Y_2 \end{smallmatrix}\right]$ and $\mathrm{rank}(\Pi_{\mathcal{N}(D)}Y\Pi_{\mathcal{N}(D)}) = \mathrm{rank}(Y_1 + Y_2)$. Consequently, (32) holds owing to [37, Thm. 9.2.4], since $Y_1 + Y_2 \succeq 0$ and Y_1 and Y_2 are parallel summable. The claim for $X^\star$ follows from Proposition 4.5(ii), since

$$\begin{aligned} X^\star &= \tfrac{1}{4}(Y_1 + Y_2) - \tfrac{1}{4}(Y_1 - Y_2)(E \otimes \mathrm{I}_n)^\top \Big((EE^\top) \otimes (Y_1 + Y_2)\Big)^\dagger (E \otimes \mathrm{I}_n)(Y_1 - Y_2) \\ &= \tfrac{1}{4}(Y_1 + Y_2) - \tfrac{1}{4}(Y_1 - Y_2)(Y_1 + Y_2)^\dagger (Y_1 - Y_2) \\ &= \tfrac{1}{2}Y_2(Y_1 + Y_2)^\dagger(Y_1 + Y_2) + \tfrac{1}{2}(Y_1 + Y_2)(Y_1 + Y_2)^\dagger Y_2 - Y_2(Y_1 + Y_2)^\dagger Y_2 \\ &= Y_1 \,\square\, Y_2, \end{aligned}$$

where the second equality holds since for arbitrary matrices Z_1, Z_2 it holds that $(Z_1 \otimes Z_2)^\dagger = Z_1^\dagger \otimes Z_2^\dagger$ and the final equality holds by definition of the parallel sum and parallel summability. □

Lemma A.3 *Suppose that Assumption III holds and that the sets* $\mathrm{gph}(\Pi_{\mathcal{R}(L^\top)}A^{-1})$ *and* $\mathrm{gph}(\Pi_{\mathcal{R}(L)}B)$ *are not singletons. Then, it holds that* $[\mu_A]_+[\rho_A]_+ \leq 1/4\sigma_d^2$ *and* $[\mu_B]_+[\rho_B]_+ \leq 1/4\sigma_d^2$.

Proof Suppose that $\mu_A, \rho_A > 0$ and $\mu_B, \rho_B > 0$, for otherwise the two claims hold trivially. Consider $y_D \in \operatorname{dom}\big(A^{-1} \circ (-L^\top)\big) \cap \operatorname{dom}\big(B^{-1}\big) = \operatorname{dom} T_{\mathrm{D}} \neq \emptyset$ and let $y = -L^\top y_D$. By semimonotonicity of A at $(x^\star, -L^\top y^\star)$ it holds for all $(x^\star, y^\star) \in \mathcal{S}^\star$, $x \in A^{-1}(y) = A^{-1} \circ (-L^\top)(y_D)$ that

$$\begin{aligned}\langle x - x^\star, y + L^\top y^\star\rangle &\geq \mathrm{q}_{\mu_A L^\top L}(x - x^\star) + \rho_A \|y + L^\top y^\star\|^2 \\ &\geq \mu_A \sigma_d^2 \|\Pi_{\mathcal{R}(L^\top)}(x - x^\star)\|^2 + \rho_A \|y + L^\top y^\star\|^2. \end{aligned} \tag{41}$$

Noting that by definition $(y, \Pi_{\mathcal{R}(L^\top)} x) \in \operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ and since $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ is not equal to the singleton $\{(-L^\top y^\star, \Pi_{\mathcal{R}(L^\top)} x^\star)\}$, both of the above involved norms are nonzero.

On the other hand, since $y = -L^\top y_D \in \mathcal{R}(L^\top)$, it holds by the Fenchel-Young inequality with modulus $2\mu_A \sigma_d^2 > 0$ that

$$\begin{aligned}\langle x - x^\star, y + L^\top y^\star\rangle &= \langle \Pi_{\mathcal{R}(L^\top)}(x - x^\star), y + L^\top y^\star\rangle \\ &\leq \mu_A \sigma_d^2 \|\Pi_{\mathcal{R}(L^\top)}(x - x^\star)\|^2 + \tfrac{1}{4\mu_A \sigma_d^2} \|y + L^\top y^\star\|^2. \end{aligned} \tag{42}$$

Combining (41) and (42), it follows that $\rho_A \leq \frac{1}{4\mu_A \sigma_d^2}$.

Analogously, consider $x_P \in \operatorname{dom}(A) \cap \operatorname{dom}(B \circ L) = \operatorname{dom} T_{\mathrm{P}} \neq \emptyset$ and let $x = Lx_P$. Then, it holds for all $(x^\star, y^\star) \in \mathcal{S}^\star$, $y \in B(x) = B(Lx_P)$ by the semimonotonicity assumption of B at $(Lx^\star, y^\star)$ that

$$\begin{aligned}\langle x - Lx^\star, y - y^\star\rangle &\geq \mu_B \|x - Lx^\star\|^2 + \mathrm{q}_{\rho_B L L^\top}(y - y^\star) \\ &\geq \mu_B \|x - Lx^\star\|^2 + \rho_B \sigma_d^2 \|\Pi_{\mathcal{R}(L)}(y - y^\star)\|^2, \end{aligned} \tag{43}$$

where the involved norms are nonzero since $(x, \Pi_{\mathcal{R}(L)} y) \in \operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ and $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ is not equal to the singleton $\{Lx^\star, \Pi_{\mathcal{R}(L)} y^\star\}$. On the other hand, since $x = Lx_P \in \mathcal{R}(L)$, it holds by the Fenchel-Young inequality with modulus $2\mu_B > 0$ that

$$\begin{aligned}\langle x - Lx^\star, y - y^\star\rangle &= \langle x - Lx^\star, \Pi_{\mathcal{R}(L)}(y - y^\star)\rangle \\ &\leq \mu_B \|x - Lx^\star\|^2 + \tfrac{1}{4\mu_B} \|\Pi_{\mathcal{R}(L)}(y - y^\star)\|^2. \end{aligned} \tag{44}$$

Finally, combining (43) and (44), it follows that $\rho_B \sigma_d^2 \leq \frac{1}{4\mu_B}$, establishing the claim. □

Proposition A.1 (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* $\left(\operatorname{diag}\left(\frac{|\tilde{v}_1|}{u_1 - l_1}, \ldots, \frac{|\tilde{v}_n|}{u_n - l_n}\right), 0\right)$*-semimonotone at* $(\tilde{x}, \tilde{v}) \in \operatorname{gph} N_C$.

Proof By Proposition 4.3(iii) it suffices to show that N_{C_i} is $\left(\frac{|\tilde{v}_i|}{u_i - l_i}, 0\right)$-semimonotone at $(\tilde{x}_i, \tilde{v}_i) \in \operatorname{gph} N_{C_i}$. Using the fact that $|\tilde{x}_i - x_i| \leq u_i - \ell_i$ and monotonicity of N_C, we have for all $x_i \in C_i$ that $\frac{|\tilde{v}_i|}{u_i - \ell_i} |\tilde{x}_i - x_i|^2 \leq |\tilde{v}_i| |\tilde{x}_i - x_i| = \langle \tilde{v}_i, \tilde{x}_i - x_i\rangle$. □

B Omitted Proofs

Proof of Lemma 3.2 (oblique weak Minty for primal and dual operator) Note that $\operatorname{gph} T_{\mathrm{PD}}$ is equal to the set $\{\big((x_A, y_B), (y_A + L^\top y_B, x_B - Lx_A)\big) | (x_A, y_A) \in \operatorname{gph} A, (x_B, y_B) \in \operatorname{gph} B\}$. Consequently, it holds by assumption for all $(x^\star, y^\star) \in \mathcal{S}^\star$, $(x_A, y_A) \in \operatorname{gph} A$ and $(x_B, y_B) \in \operatorname{gph} B$ that

$$\begin{aligned} &\left\langle y_A + L^\top y_B, x_A - x^\star \right\rangle + \left\langle x_B - Lx_A, y_B - y^\star \right\rangle \\ &\quad \geq \mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B) + \mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A). \end{aligned} \tag{45}$$

- In (45), consider $x_A \in \operatorname{dom}(A) \cap \operatorname{dom}(B \circ L) = \operatorname{dom} T_{\mathrm{P}} \neq \emptyset$ and let $x_B = Lx_A$. Then, it holds for all $x^\star \in \mathcal{S}^\star_{\mathrm{P}}$, $y_A \in A(x_A)$ and $y_B \in B(Lx_A)$ that $\left\langle y_A + L^\top y_B, x_A - x^\star \right\rangle \geq \mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B)$. Since $(x_A, y_A + L^\top y_B) \in \operatorname{gph} T_{\mathrm{P}}$ by construction and $\mathcal{S}^\star_{\mathrm{P}} \subseteq \operatorname{zer} T_{\mathrm{P}}$ by Proposition 3.1, it follows by definition that T_{P} has V_{P}-oblique weak Minty solutions at $\mathcal{S}^\star_{\mathrm{P}}$.
- Analogously, consider $y_B \in \operatorname{dom}\big(A^{-1} \circ (-L^\top)\big) \cap \operatorname{dom}\big(B^{-1}\big) = \operatorname{dom} T_{\mathrm{D}} \neq \emptyset$ and let $y_A = -L^\top y_B$ in (45). Then, it holds for all $y^\star \in \mathcal{S}^\star_{\mathrm{D}}$, $x_A \in A^{-1} \circ (-L^\top)(y_B)$ and $x_B \in B^{-1}(y_B)$ that $\langle x_B - Lx_A, y_B - y^\star \rangle \geq \mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A)$. Since $(y_B, x_B - Lx_A) \in \operatorname{gph} T_{\mathrm{D}}$ by construction and $\mathcal{S}^\star_{\mathrm{D}} \subseteq \operatorname{zer} T_{\mathrm{D}}$ by Proposition 3.1, it follows by definition that T_{D} has V_{D}-oblique weak Minty solutions at $\mathcal{S}^\star_{\mathrm{D}}$, completing the proof.

□

Proof of Example 3.1 (saddle point problem) ♠ 3.1(i): By defining $H := \mathrm{I} + \lambda\big((P + T_{\mathrm{PD}})^{-1} P - I\big)$ and substituting $\tau = 1/\gamma \ell^2$, the update rule for z^k corresponds to the linear dynamical system $z^{k+1} = Hz^k$. Global asymptotic stability of this system is achieved if and only if the spectral radius of the matrix H is strictly less than one, which holds iff $\lambda \in \big(0, \min\big\{2, \bar{\lambda}\big\}\big)$. Analogously, the convergence result for $(\Pi_{\mathcal{R}(P)} z^k)_{k \in \mathbb{N}}$ can be obtained by analyzing the spectral radius of $\Pi_{\mathcal{R}(P)} H$.

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

$$\begin{aligned} T_{\mathrm{PD}} &= \begin{bmatrix} A & L^\top \\ -L & B^{-1} \end{bmatrix} \quad \text{and} \\ T_{\mathrm{PD}}^{-1} &= \begin{bmatrix} (A + L^\top BL)^{-1} & -(A + L^\top BL)^{-1} L^\top B \\ BL(A + L^\top BL)^{-1} & B - BL(A + L^\top BL)^{-1} L^\top B \end{bmatrix} \end{aligned} \tag{46}$$

owing to the Schur complement lemma. Therefore, T_{PD} has a V-oblique weak Minty solution at zero on $\mathcal{U} = \mathbb{R}^5 \times \mathcal{R}(P)$ if and only if

$$z^\top \left(\tfrac{T_{\mathrm{PD}} + T_{\mathrm{PD}}^\top}{2} - T_{\mathrm{PD}}^\top V T_{\mathrm{PD}} \right) z \geq 0, \qquad \text{for all } z \in \mathbb{R}^n : z \in T_{\mathrm{PD}}^{-1} \mathcal{R}(P). \tag{47}$$

Using that $L = Y \Sigma X^\top$, where $Y = \mathrm{I}_3$, $\Sigma = \left[\begin{smallmatrix} |\ell| \mathrm{I}_2 \\ 0 \end{smallmatrix}\right]$ and $X = \operatorname{sgn}(\ell) \mathrm{I}_2$, it follows by plugging in (15) into (16) that

$$U = \begin{bmatrix} \frac{1}{\sqrt{1+\gamma^2\ell^2}}\mathrm{I}_2 & 0 \\ -\frac{\gamma\ell}{\sqrt{1+\gamma^2\ell^2}}\mathrm{I}_2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad V = \tfrac{b\ell^2}{a^2+b^2\ell^4}\mathrm{I}_2 \oplus \tfrac{ba^2}{a^2+b^2\ell^4}\mathrm{I}_2 \oplus c\mathrm{I}_1, \tag{48}$$

where U is an orthonormal basis for $\mathcal{R}(P)$. As a result, (47) is satisfied if and only if

$$\left(T_{\mathrm{PD}}^{-1}U\right)^\top \left(\tfrac{T_{\mathrm{PD}}+T_{\mathrm{PD}}^\top}{2} - T_{\mathrm{PD}}^\top V T_{\mathrm{PD}}\right) T_{\mathrm{PD}}^{-1}U \succeq 0.$$

Noting that the left-hand side becomes zero when plugging in (46) and (48), the claim follows.

- ♠ 3.1(iii): Since condition (10) holds for $\beta_{\mathrm{P}} = \frac{b\ell^2}{a^2+b^2\ell^4}$, $\beta_{\mathrm{D}} = \frac{ba^2}{a^2+b^2\ell^4}$ and $\beta'_{\mathrm{D}} = c$, it follows from Example 3.1(ii) that Assumption II.A3 is satisfied. Moreover, it can be verified that for these parameters δ is equal to one, that Stepsize rule I reduces to (31) and that Relaxation parameter rule I matches (30). Operators A and B are continuous and hence outer semicontinuous, verifying Assumption II.A1. It only remains to show that $J_{\gamma A}$ and $J_{\tau B^{-1}}$ have full domain and are single-valued continuous. For $J_{\gamma A}$ this holds since $\mathrm{I}+\gamma A$ is invertible for any $\gamma \in \mathbb{R}$. For $J_{\tau B^{-1}}$ with $\tau = 1/\gamma\ell^2$, invertability of $\mathrm{I} + \tau B^{-1}$ is ensured provided that $-\gamma b\ell^2 \neq 1$ and $-\gamma c\ell^2 \neq 1$. Since these conditions are guaranteed by Stepsize rule I and the assumption $c \geq 0$, the claim is established.
- ♠ 3.1(iv): For this particular instance $\operatorname{tr} T_{\mathrm{P}} = -2$ and $\operatorname{tr} T_{\mathrm{D}} = \operatorname{tr} T_{\mathrm{PD}} = -8$. Since the trace of a matrix equals the sum of its eigenvalues, the proof is completed.

□

Proof of Example 3.2 (influence of singular values) ♠ 3.2(i): By [6, Prop. 5.1(ii)] and using that A, B and L are symmetric, T_{PD} is $\frac{1}{2}$-comonotone if and only if

$$\tfrac{T_{\mathrm{PD}}+T_{\mathrm{PD}}^\top}{2} - \tfrac{1}{2}T_{\mathrm{PD}}^\top T_{\mathrm{PD}} = \begin{bmatrix} A - \frac{1}{2}(A^\top A + L^\top L) & -\frac{1}{2}(AL - LB^{-1}) \\ -\frac{1}{2}(LA - B^{-1}L) & B^{-1} - \frac{1}{2}({B^{-1}}^\top B^{-1} + LL^\top) \end{bmatrix} \tag{49}$$

is positive semidefinite. Using that $B^{-1} = A$, that L is symmetric and that A and L commute, i.e., $AL = LA$, this condition reduces to $A - \frac{1}{2}(A^\top A + L^\top L) \succeq 0$, which holds by definition of A and L. Noting that $\operatorname{zer} T_{\mathrm{PD}} = (0_n, 0_n)$, the claim is established.

- ♠ 3.2(ii): Follows from Theorem 3.1 and Proposition 3.2, using that $\|L\| = 1$ and $\gamma\tau = \frac{1}{\|L\|^2}$.
- ♠ 3.2(iii): Analogous to the setting of Example 3.1(i), the update rule for $\Pi_{\mathcal{R}(P)}z^k$ can be expressed as the linear dynamical system $\Pi_{\mathcal{R}(P)}z^{k+1} = \Pi_{\mathcal{R}(P)}H\Pi_{\mathcal{R}(P)}z^k$, where $H{:=}\mathrm{I} + \lambda\left((P+T_{\mathrm{PD}})^{-1}P - I\right)$ and P is defined as in (1). This system is globally asymptotically stable if and only if the spectral radius of $\Pi_{\mathcal{R}(P)}H$ is

strictly less than one, i.e., if and only if $\lambda \in (0, \bar{\lambda}_{\text{spectral}})$, where

$$\bar{\lambda}_{\text{spectral}} \in \arg\max_{\lambda} \ \lambda \ \text{ subject to } \ \|\Pi_{\mathcal{R}(P)} H(\lambda)\|_2 < 1. \tag{50}$$

The values for $\bar{\lambda}_{\text{spectral}}$ reported in Fig. 4 are obtained by solving this problem using SymPy.

□

Proof of Proposition 4.1 By the Fenchel-Young inequality, it holds for any $R \prec 0$ that $\langle x - x', y - y' \rangle \geq \frac{1}{4} \mathrm{q}_{R^{-1}}(x - x') - \mathrm{q}_R(y - y')$, for all $(x, y), (x', y') \in \operatorname{gph} A$. Therefore, (3) is satisfied for all $M \preceq \frac{1}{4} R^{-1}$. □

Proof of Proposition 4.2 (preservation of maximality) If T is maximally (M', R')-semimonotone, then by definition there exists no (M', R')-semimonotone operator $\tilde{T}$ such that $\operatorname{gph} T \subset \operatorname{gph} \tilde{T}$. Since the class of (M', R')-semimonotone operator encompasses the class of (M, R)-semimonotone operators whenever $M' \preceq M$ and $R' \preceq R$, this implies that there also exists no (M, R)-semimonotone operator $\tilde{T}$ such that $\operatorname{gph} T \subset \operatorname{gph} \tilde{T}$, establishing the claim by definition of maximality. □

Proof of Proposition 4.3 (inverting, shifting, scaling and cartesian product) ♠ 4.3(ii): First, consider the assertion where A is semimonotone only at (x'_A, y'_A). Define $s' = x'_A - u$ and $t' = y + \alpha y'_A$, such that $(s', t') \in \operatorname{gph} T$. Then, it holds for all $(s, t) \in \operatorname{gph} T$ that

$$\begin{aligned}
\langle s - s', t - t' \rangle &= \alpha \left\langle (s + u) - (s' + u), \alpha^{-1}(t - y) - \alpha^{-1}(t' - y) \right\rangle \\
&\geq \alpha \, \mathrm{q}_M((s + u) - (s' + u)) \\
&\quad + \alpha \, \mathrm{q}_R(\alpha^{-1}(t - y) - \alpha^{-1}(t' - y)) \\
&= \mathrm{q}_{\alpha M}(s - s') + \mathrm{q}_{\alpha^{-1} R}(t - t'),
\end{aligned}$$

where we used that $(s' + u, \alpha^{-1}(t' - y)) = (x'_A, y'_A)$, that $(s + u, \alpha^{-1}(t - y)) \in \operatorname{gph} A$ and semimonotonicity of A at(x'_A, y'_A). Hence, it follows that T is $(\alpha M_A, \alpha^{-1} R_A)$-semimonotone at (s', t').

If A is (M_A, R_A)-semimonotone at all $(x'_A, y'_A) \in \operatorname{gph} A$, we then know that T is $(\alpha M_A, \alpha^{-1} R_A)$-semimonotone at all points in the set $\{(x'_A - u, y + \alpha y'_A) | (x'_A, y'_A) \in \operatorname{gph} A\}$. Since this set is equal to $\operatorname{gph} T$, it follows that T is $(\alpha M_A, \alpha^{-1} R_A)$-semimonotone (everywhere).

♠ 4.3(iii): Let A and B be semimonotone at respectively (x'_A, y'_A) and (x'_B, y'_B). Since $\operatorname{gph} T$ is equal to the set $\{((x_A, x_B), (y_A, y_B)) | y_A \in A x_A, y_B \in B x_B\}$, it holds for all $(x, y) \in \operatorname{gph} T$ that

$$\begin{aligned}
\langle x - x', y - y' \rangle &= \langle x_A - x'_A, y_A - y'_A \rangle + \langle x_B - x'_B, y_B - y'_B \rangle \\
&\geq \mathrm{q}_{M_A}(x_A - x'_A) + \mathrm{q}_{R_A}(y_A - y'_A) \\
&\quad + \mathrm{q}_{M_B}(x_B - x'_B) + \mathrm{q}_{R_B}(y_B - y'_B) \\
&= \mathrm{q}_{M_A \oplus M_B}(x - x') + \mathrm{q}_{R_A \oplus R_B}(y - y'),
\end{aligned}$$

where the semimonotonicity of A and B was used in the inequality, and thus T is $\big(M_A \oplus M_B,\, R_A \oplus R_B\big)$-semimonotone at $(x', y') \in \operatorname{gph} T$.
If A and B are semimonotone at all points in their graph, then T is $\big(M_A \oplus M_B,\, R_A \oplus R_B\big)$-semimonotone at all points in $\operatorname{gph} T$, which completes the proof.

□

Proof of Proposition 4.4 (linear operator) Owing to the linearity of D, (M, R)-semimonotonicity corresponds to having $\langle x, Dx\rangle \geq \mathrm{q}_M(x) + \mathrm{q}_R(Dx)$ for all $x \in \mathbb{R}^n$, which is equivalent to the claimed LMI. □

Proof of Proposition 4.5 (symmetric solution of $D^\top X D \preceq Y$) ♠ 4.5(i): First, note that the problem of finding an $X \in \mathbb{S}^n$ such that $D^\top X D \preceq Y$ is equivalent to the problem of finding a pair $(X, Z) \in \mathbb{S}^n \times \mathbb{S}^m$ such that $D^\top X D = Y - Z$ and $Z \succeq 0$. Second, observe that by [42, Thm. 2], [24, Prop. 1] the involved linear matrix equality is solvable for $X \in \mathbb{S}^n$ if and only if $\mathcal{R}(Y - Z) \subseteq \mathcal{R}(D^\top)$, i.e., $\Pi_{\mathcal{N}(D)} Z = \Pi_{\mathcal{N}(D)} Y$. By [29, Thm. 2.2], a matrix $Z \succeq 0$ satisfying this condition exists if and only if (32) holds, and the general solution is given by

$$Z = Y\Pi_{\mathcal{N}(D)}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y + D^\top (D^\dagger)^\top G D^\top (D^\dagger)^\top, \tag{51}$$

where $G \in \mathbb{S}^n$ is an arbitrary symmetric positive semidefinite matrix.
♠ 4.5(ii): Using (51), the general solution of $D^\top X D = Y - Z$ is given by [42, Thm. 2], [24, Prop. 1]

$$\begin{aligned} X = \underbrace{(D^\dagger)^\top (Y - Y\Pi_{\mathcal{N}(D)}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y - G) D^\dagger}_{= X^\star - (D^\dagger)^\top G D^\dagger} \\ + H - \Pi_{\mathcal{R}(D)} H \Pi_{\mathcal{R}(D)}, \end{aligned} \tag{52}$$

where $H \in \mathbb{S}^n$ is an arbitrary matrix. Substituting (52) into $D^\top X D \preceq Y$ shows that $Y - D^\top X D = Y - D^\top X^\star D + \Pi_{\mathcal{R}(D^\top)} G \Pi_{\mathcal{R}(D^\top)} \succeq Y - D^\top X^\star D \succeq 0$. Finally, the alternative expression for $X^\star$ given in (33) follows directly from [58, Lem. 3].
♠ 4.5(iii): [7, Fact 6.4.38]

□

Proof of Corollary 4.2 (semimonotonicity of $DTD^\top$) If (32) holds, then it follows from Proposition 4.5 that $D^\top X^\star D \preceq Y$. Therefore, it only remains to be to shown that this implies $(DMD^\top, X^\star)$-semimonotonicity of $DTD^\top$ [at (x', Dy')]. First, consider the case where T is semimonotone only at a single point $(D^\top x', y')$. Let $(D^\top x, y) \in \operatorname{gph} T$ and denote $u = Dy$ and $u' = Dy'$. Then, $(x, u), (x', u') \in \operatorname{gph} DTD^\top$ and it holds that

$$\begin{aligned} \langle x - x', u - u'\rangle &= \langle D^\top (x - x'), y - y'\rangle \\ (\text{semi. of } T \text{ at } (D^\top x', y')) &\geq \mathrm{q}_M(D^\top (x - x')) + \mathrm{q}_Y(y - y') \\ &\geq \mathrm{q}_{DMD^\top}(x - x') + \mathrm{q}_{X^\star}(D(y - y')) \end{aligned}$$

$$= \mathrm{q}_{DMD^\top}(x - x') + \mathrm{q}_{X^\star}(u - u'),$$

where $D^\top X^\star D \preceq Y$ was used in the second inequality, showing that $DTD^\top$ is $(DMD^\top, X^\star)$-semimonotone at (x', Dy'). Finally, if T is (M, Y)-semimonotone at all $(D^\top x', y') \in \operatorname{gph} T$, then $DTD^\top$ is $(DMD^\top, X^\star)$-semimonotone at all points in $\{(x', Dy') | (D^\top x', y') \in \operatorname{gph} T\}$, which equals $\operatorname{gph} DTD^\top$. □

Proof of Proposition 4.6 (sum and parallel sum). Let $D = [\,\mathrm{I}_n\ \mathrm{I}_n\,]$. Then, $A + B$ is equal to $DTD^\top$, where $T := A \times B$. By Proposition 4.3(iii), operator T is $(M, R) = \big(M_A \oplus M_B, R_A \oplus R_B\big)$-semimonotone [at $((x', x'), (y'_A, y'_B)) \in \operatorname{gph} T$]. Consequently, it follows from Corollary 4.2 and Lemma A.2 that $DTD^\top = A + B$ is $(M_A + M_B, R_A \,\square\, R_B)$-semimonotone [at $(x', y'_A + y'_B)$]. Finally, the claim for the parallel sum follows directly from those for the sum and Proposition 4.3(i), since $A \,\square\, B := (A^{-1} + B^{-1})^{-1}$. □

Proof of Lemma 4.1 (sum with skew-symmetric matrix) First, consider the assertion where T is semimonotone at (x', y'). Let $(x, y) \in \operatorname{gph} T$. Then, $(x, y + Dx)$, $(x', y' + Dx') \in \operatorname{gph}(T + D)$ and $\langle x - x', y - y' + D(x - x')\rangle \geq \mathrm{q}_{D^\top M D}(x - x') + \mathrm{q}_{R+R'}(y - y') + \langle x - x', D(x - x')\rangle$. By Corollary 4.1(ii) and skew-symmetry of D, it follows that D is $(-D^\top M D, \Pi_{\mathcal{R}(D)} M \Pi_{\mathcal{R}(D)})$-semimonotone. Consequently,

$$\begin{aligned}
&\langle x - x', y - y' + D(x - x')\rangle \\
&\quad \geq \mathrm{q}_{R+R'}(y - y') + \mathrm{q}_{\Pi_{\mathcal{R}(D)} M \Pi_{\mathcal{R}(D)}}(Dx - Dx') \\
&\quad = \mathrm{q}_{R'}(y - y' + D(x - x')) + \mathrm{q}_{R}(y - y') + \mathrm{q}_{M}(D(x - x')) \\
&\quad = \mathrm{q}_{R'}(y - y' + D(x - x')) + q_{M \oplus R}\left((D(x - x'), y - y')\right), \qquad (53)
\end{aligned}$$

where the fact that $\mathcal{R}(R') \subseteq \mathcal{N}(D)$ was used in the first equality. Let $E := [\,\mathrm{I}_n\ \mathrm{I}_n\,]$. Since $(M, R) \in \operatorname{dom}_{\square}$, it holds by Proposition 4.5 and Lemma A.2 that $E^\top (M \,\square\, R) E \preceq M \oplus R$. Consequently, it follows from (53) that

$$\langle x - x', y - y' + D(x - x')\rangle \geq \mathrm{q}_{R' + M \square R}(y - y' + D(x - x')),$$

proving the claimed semimonotonicity of $T + D$ at $(x', y' + Dx')$. Finally, when T is semimonotone everywhere, the claim follows analogously by considering all $(x', y') \in \operatorname{gph} T$, completing the proof. □

Proof of Proposition 5.1 (constrained QP) ♠ 5.1(i): Let $M := L^\top M_A L = \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)}$ and observe that

$$\Pi_{\mathcal{N}(Q)} M = \Pi_{\mathcal{N}(Q)} \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)} = \Pi_{\mathcal{N}(Q)} (\mathrm{I}_n - \Pi_{\mathcal{N}(L)}) Q \Pi_{\mathcal{R}(L^\top)} = 0, \qquad (54)$$

where the final equality holds since $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$. Therefore, (34) is satisfied for $D = Q$ and Q is $(L^\top M_A L, R^\star)$-semimonotone by Corollary 4.1(ii), where $R^\star = Q^\dagger - Q^\dagger M Q^\dagger = Q^\dagger - Q^\dagger \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)} Q^\dagger$ due to (54) and symmetry

of Q. Moreover, since $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$, it holds by [7, Fact 6.4.34] that $Q^\dagger = (\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)})^\dagger + (\Pi_{\mathcal{N}(L)} Q \Pi_{\mathcal{N}(L)})^\dagger$, so that $R^\star = \Pi_{\mathcal{N}(L)} Q^\dagger \Pi_{\mathcal{N}(L)} = R'_A$. The claimed result for $A : x \mapsto Qx + q$ then follows from Proposition 4.3(ii).

- ♠ 5.1(ii): By Proposition 3.1, it holds that $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ if and only if $(x^\star, -L^\top y^\star) \in \operatorname{gph} A$ and $(Lx^\star, y^\star) \in \operatorname{gph} B$. The claimed result then follows directly from Proposition A.1.
- ♠ 5.1(iii): By Theorem 5.1, T_{PD} has $((\Pi_{\mathcal{N}(L)} Q^\dagger \Pi_{\mathcal{N}(L)}) \oplus (M_A \square M_B))$-oblique weak Minty solutions at $\mathcal{S}^\star = \{(x^\star, y^\star)\}$. Therefore, recalling that $\Pi_{\mathcal{N}(L)} = X' X'^\top$, it follows that the parameters β_{P}, β_{D}, β'_{P} and β'_{D} given in 5.1(iii) match those from (37). As remarked below Theorem 5.1, Assumption II.A3 holds since these parameters satisfy condition (10), completing the proof.
- ♠ 5.1(iv): Note that both A and B are outer semicontinuous since A is a linear mapping and B is a normal cone [50, Prop. 6.6], verifying Assumption II.A1. Since the parameters β_{P}, β_{D}, β'_{P} and β'_{D} from Example 5.1(iii) satisfy condition (10), it follows that also Assumption II.A3 holds. Moreover, it is straightforward to verify that for these parameter values Stepsize rule I and Relaxation parameter rule I correspond to (39) (excluding the case $\tau = 1/\gamma \|L\|^2$ for simplicity). Therefore, it only remains to show that the resolvents $J_{\gamma A}$ and $J_{\tau^{-1} B}$ have full domain and are (single-valued) continuous for any γ and τ complying with (39). Note that for any $\tau > 0$ it holds that $J_{\tau B^{-1}} = \mathrm{id} - \tau J_{\tau^{-1} B} \circ (\tau^{-1} \mathrm{id})$. Since $J_{\tau^{-1} B} = \Pi_C$ is (single-valued) continuous and has full domain, this implies that so does $J_{\tau B^{-1}}$. On the other hand, since $A(x) = Qx + q$ it holds that $J_{\gamma A}(\cdot) = J_{\gamma Q}(\cdot - \gamma q)$. Therefore, $J_{\gamma A}$ has full domain and is (single-valued) continuous if and only if so is $J_{\gamma Q}$, i.e., if and only if the matrix $\mathrm{id} + \gamma Q$ is invertible. We distinguish between two cases.

  - $\lambda_{\min}(Q) \geq 0$: Then, $\mathrm{id} + \gamma Q$ is invertible for any $\gamma > 0$ and the claim is immediate.
  - $\lambda_{\min}(Q) < 0$: Let X, X', Y, Y' be defined as in (6), so that $L = Y \Sigma X^\top$. We begin by showing the following intermediate claims.

Claim 1 *The smallest eigenvalue of Q is equal to the smallest eigenvalue of $X^\top Q X$.*

Since by assumption $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$ where $\Pi_{\mathcal{R}(L^\top)} = X X^\top$ and $\Pi_{\mathcal{N}(L)} = X' X'^\top$, we have

$$Q = \begin{bmatrix} X & X' \end{bmatrix} \underbrace{\begin{bmatrix} X^\top Q X & 0 \\ 0 & X'^\top Q X' \end{bmatrix}}_{=: \bar{Q}} \begin{bmatrix} X^\top \\ X'^\top \end{bmatrix}.$$

Consequently, it follows by the similarity transformation with similarity matrix $\begin{bmatrix} X & X' \end{bmatrix}$ [27, Cor. 1.3.4] that Q is similar to $\bar{Q}$, and therefore $\lambda_{\min}(Q) = \min\{\lambda_{\min}(X^\top Q X), \lambda_{\min}(X'^\top Q X')\}$, where we used the block diagonal structure of $\bar{Q}$. Moreover, since we assumed that $\Pi_{\mathcal{N}(L)} Q \Pi_{\mathcal{N}(L)} \succeq 0$ it follows that also $X'^\top Q X' = X'^\top \Pi_{\mathcal{N}(L)} Q \Pi_{\mathcal{N}(L)} X' \succeq 0$ and since we assumed that $\lambda_{\min}(Q) < 0$ necessarily $\lambda_{\min}(Q) = \lambda_{\min}(X^\top Q X)$.

Claim 2 *The parameter β_{D} is upper bounded by $\frac{1}{\|L\|^2}\lambda_{\min}(Q)$.*

Since $(M_A, M_B(y^\star)) \in \mathrm{dom}\square$, we have that $M_A + M_B(y^\star) \succeq 0$ and consequently it follows from (35) that $M_A \square M_B(y^\star) \preceq M_A$, where $M_A = {L^\dagger}^\top Q L^\dagger$. Since $L^\dagger = X\Sigma^{-1}Y^\top$ and $Y^\top Y = \mathrm{I}$, this implies that

$$\begin{aligned}\beta_{\mathrm{D}} &= \lambda_{\min}\left(Y^\top\left(M_A \square M_B(y^\star)\right)Y\right)\\ &\leq \lambda_{\min}\left(\Sigma^{-1}X^\top Q X \Sigma^{-1}\right)\\ &= \min_{s\in\mathbb{R}^r\setminus\{0\}} \frac{s^\top\Sigma^{-1}X^\top Q X\Sigma^{-1}s}{s^\top s}\\ &= \min_{t\in\mathbb{R}^r\setminus\{0\}} \frac{t^\top X^\top Q X t}{t^\top\Sigma^2 t},\end{aligned}$$

where we defined $t{:=}\Sigma^{-1}s$ in the final step. Since $\lambda_{\min}(X^\top Q X) < 0$ due to Claim 1 and since $\Sigma^2 \preceq \|L\|^2\mathrm{I}$, it follows that

$$\beta_{\mathrm{D}} \leq \tfrac{1}{\|L\|^2}\min_{t\in\mathbb{R}^r\setminus\{0\}}\frac{t^\top X^\top Q X t}{t^\top t} = \tfrac{1}{\|L\|^2}\lambda_{\min}\left(X^\top Q X\right) = \tfrac{1}{\|L\|^2}\lambda_{\min}\left(Q\right).$$

Finally, it follows from Claim 2 that $\mathrm{I}+\gamma Q \succeq \big(1+\gamma\lambda_{\min}(Q)\big)\mathrm{I} \succeq \big(1+\gamma\beta_{\mathrm{D}}\|L\|^2\big)\mathrm{I}$, showing that $\mathrm{I}+\gamma Q$ is invertible since $\gamma < 1/[\beta_{\mathrm{D}}]_-\|L\|^2$ as in (39). □

Proof of Example 5.1 The projections are given by $\Pi_{\mathcal{R}(L^\top)} = \mathrm{diag}(1,1,0)$ and $\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,1)$, leading to $\Pi_{\mathcal{R}(L^\top)}Q\Pi_{\mathcal{N}(L)} = 0$ and $\Pi_{\mathcal{N}(L)}Q\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,2) \succeq 0$. Therefore, we can invoke Proposition 5.1 to establish the assertions. In doing so, the following intermediate computations arise: $L^\dagger = \left[\begin{smallmatrix}1 & 0 & 0\\ -\frac{1}{4} & 1 & 0\end{smallmatrix}\right]^\top$, $\|L\| = \frac{1}{8}(1+\sqrt{65})$, $M_A = {L^\dagger}^\top Q L^\dagger = \frac{1}{16}\left[\begin{smallmatrix}16 & -4\\ -4 & -15\end{smallmatrix}\right]$, $L^\top M_A L = \mathrm{diag}(1,-1,0)$, $R'_A = \Pi_{\mathcal{N}(L)}Q^{-1}\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,1/2)$, $M_B = \mathrm{diag}(0,3/2)$, $X' = [0\,0\,1]^\top$, ${X'}^\top Q^{-1}X' = 1/2$, $M_A \square M_B = \mathrm{diag}(0,-3)$ and that $\beta_{\mathrm{D}} = \lambda_{\min}\left(Y^\top(\mathrm{diag}(0,-3))\,Y\right) = -3$ since Y is an orthonormal basis with full rank. □

Proof of Example 5.2 The proof is analogous to the one of Example 5.1. It holds that $\Pi_{\mathcal{R}(L^\top)} = \mathrm{diag}(1,1,0)$ and $\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,1)$. Therefore, $\Pi_{\mathcal{R}(L^\top)}Q\Pi_{\mathcal{N}(L)} = 0$ and $\Pi_{\mathcal{N}(L)}Q\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,1) \succeq 0$. Hence, we can invoke Proposition 5.1 to establish the assertions. In doing so, the following intermediate computations arise: $L^\dagger = \frac{1}{6}\left[\begin{smallmatrix}2 & 2 & 2\\ 0 & 3 & -3\\ 0 & 0 & 0\end{smallmatrix}\right]$, $\|L\| = \sqrt{3}$, $M_A = {L^\dagger}^\top Q L^\dagger = -\mathrm{I}_3$, $R'_A = \Pi_{\mathcal{N}(L)}Q^{-1}\Pi_{\mathcal{N}(L)} = \mathrm{diag}(0,0,1)$, $M_B = 2\mathrm{I}_3$, $X' = [0\,0\,1]^\top$, $Y = \left[\begin{smallmatrix}1/\sqrt{3} & 0\\ 1/\sqrt{3} & 1/\sqrt{2}\\ 1/\sqrt{3} & -1/\sqrt{2}\end{smallmatrix}\right]$, ${X'}^\top Q^{-1}X' = 1$ and $M_A \square M_B = -2\mathrm{I}_3$. □

Proof of Corollary 5.2 (convergence of CPA under semimonotonicity) To show that the claims from Theorem 3.1 hold, it suffices to verify that Assumption II.A3, Stepsize

rule I and Relaxation parameter rule I hold. By Corollary 5.1, the primal-dual operator T_{PD} has V-oblique weak Minty solutions at $\mathcal{S}^\star$, where V is defined as in (8) with parameters β_{P}, β'_{P}, β_{D} and β'_{D} given by

$$\beta_{\mathrm{P}} = \rho_A \,\square\, \rho_B, \quad \beta_{\mathrm{D}} = \mu_A \,\square\, \mu_B, \quad \beta'_{\mathrm{P}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = n, \\ \rho_A, & \text{if } \operatorname{rank} L < n, \end{cases},$$
$$\beta'_{\mathrm{D}} = \begin{cases} \infty, & \text{if } \operatorname{rank} L = m, \\ \mu_B, & \text{if } \operatorname{rank} L < m. \end{cases} \tag{55}$$

To show that Assumption II.A3 holds, we need to verify that these parameters satisfy (10), i.e., that $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < 1/4\|L\|^2$ and that

$$\beta'_{\mathrm{P}} \geq \begin{cases} 0, & \text{if } \beta_{\mathrm{P}} \geq 0, \\ \frac{\beta_{\mathrm{P}}}{1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}, & \text{if } \beta_{\mathrm{P}} < 0, \beta_{\mathrm{D}} < 0, \\ \frac{\beta_{\mathrm{P}}}{1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2}, & \text{if } \beta_{\mathrm{P}} < 0, \beta_{\mathrm{D}} \geq 0, \end{cases} \tag{56}$$

$$\beta'_{\mathrm{D}} \geq \begin{cases} 0, & \text{if } \beta_{\mathrm{D}} \geq 0, \\ \frac{\beta_{\mathrm{D}}}{1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}, & \text{if } \beta_{\mathrm{D}} < 0, \beta_{\mathrm{P}} < 0, \\ \frac{\beta_{\mathrm{D}}}{1-\beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2}, & \text{if } \beta_{\mathrm{D}} < 0, \beta_{\mathrm{P}} \geq 0. \end{cases} \tag{57}$$

The first inequality holds owing to (55) and Assumption III.(iv). Before verifying (56) and (57), we first show three intermediate results.

Claim 1 *Condition (56) holds for all $\beta'_{\mathrm{P}} \geq 0$ and (57) holds for all $\beta'_{\mathrm{D}} \geq 0$.*

We already established that $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < 1/4\|L\|^2$ and therefore $1 - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2 > 3/4\|L\|^2$ when $\beta_{\mathrm{P}} < 0$ and $\beta_{\mathrm{D}} < 0$. Moreover, it holds by construction that $1 - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma_d^2 \geq 1$ when $\beta_{\mathrm{P}}\beta_{\mathrm{D}} \leq 0$. Consequently, the fractions appearing in the cases of (56) and (57) are strictly negative as their numerators are strictly positive and their denominators strictly negative, proving the claim.

Claim 2 *It holds that $\beta_{\mathrm{P}} = \rho_A \,\square\, \rho_B \leq \rho_A$ and $\beta_{\mathrm{D}} = \mu_A \,\square\, \mu_B \leq \mu_B$.*

This claim follows by definition of the parallel sum, since either $\rho_A = \rho_B = 0$ or $\rho_A + \rho_B > 0$ and either $\mu_A = \mu_B = 0$ or $\mu_A + \mu_B > 0$ owing to Assumption III.

Claim 3 *The graphs $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} \circ A^{-1})$ and $\operatorname{gph}(\Pi_{\mathcal{R}(L)} \circ B)$ are not singletons.*

By Assumption II.A2 the resolvents $J_{\gamma A}$ and $J_{\tau B^{-1}}$ have full domain. Consequently, $\gamma\mathrm{id} + A^{-1}$ and $\tau\mathrm{id} + B$ have full range [6, Prop. 2.11(ii)], which implies that $\mathcal{R}(\Pi_{\mathcal{R}(L^\top)} \circ (\gamma\mathrm{id} + A^{-1})) = \mathcal{R}(L^\top)$ and $\mathcal{R}(\Pi_{\mathcal{R}(L)} \circ (\tau\mathrm{id} + B)) = \mathcal{R}(L)$. Since $\mathcal{R}(\Pi_{\mathcal{R}(L^\top)} \circ (\gamma\mathrm{id} + A^{-1}))$ is a singleton whenever $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} \circ A^{-1})$ is a singleton

and $\mathcal{R}(\Pi_{\mathcal{R}(L)} \circ (\tau \mathrm{id} + B))$ is a singleton whenever $\mathrm{gph}(\Pi_{\mathcal{R}(L)} \circ B)$ is a singleton, the claim is established.

Using these claims, we will now show that condition (56) holds. For any $\beta'_{\mathrm{P}} \geq 0$ this follows directly from Claim 1. Consider the case where $\beta'_{\mathrm{P}} < 0$, which by (55) implies that rank $L < n$, $\beta'_{\mathrm{P}} = \rho_A$ and $\rho_A < 0$. Then, it follows from Assumption III that (i) $\rho_B > -\rho_A > 0$, which implies by definition of β_{P} from (55) that $\beta_{\mathrm{P}} = \rho_A \square \rho_B < 0$, and that (ii) either $\mu_A = \mu_B = 0$ or $\mu_A + \mu_B > 0$.

- ♠ $\mu_A = \mu_B = 0$: By definition of β_{D} from (55) it then follows that $\beta_{\mathrm{D}} = \mu_A \square \mu_B = 0$ and that condition (56) reduces to $\rho_A \geq \beta_{\mathrm{P}}$, which holds by Claim 2.
- ♠ $\mu_A + \mu_B > 0$: Then, it holds that $\beta_{\mathrm{D}} \neq 0$, and since $\beta_{\mathrm{P}} < 0$ condition (56) reduces to

$$\rho_A \geq \frac{\beta_{\mathrm{P}}}{1 - \beta_{\mathrm{P}}([\beta_{\mathrm{D}}]_+ \sigma_d^2 - [\beta_{\mathrm{D}}]_- \|L\|^2)}.$$

Plugging in the definitions from (55), this can be equivalently stated as

$$\begin{aligned} \rho_A &\geq \frac{\rho_A \square \rho_B}{1 - (\rho_A \square \rho_B)\Big([\mu_A \square \mu_B]_+ \sigma_d^2 - [\mu_A \square \mu_B]_- \|L\|^2\Big)} \\ &= \frac{\rho_A \rho_B (\mu_A + \mu_B)}{\rho_A + \rho_B - \rho_A \rho_B \Big([\mu_A \mu_B]_+ \sigma_d^2 - [\mu_A \mu_B]_- \|L\|^2\Big)}, \end{aligned}$$

where we multiplied the numerator and denominator by $(\mu_A + \mu_B)(\rho_A + \rho_B) > 0$ in the equality. Multiplying both sides of the inequality by the strictly positive denominator (see the proof of Claim 1) and reordering, we obtain $\rho_A^2 \left(\mu_A + \mu_B - [\mu_A \mu_B]_+ \rho_B \sigma_d^2 + [\mu_A \mu_B]_- \rho_B \|L\|^2\right) \geq 0$, which is vacuously satisfied when $[\mu_A \mu_B]_+ = 0$. Otherwise, it holds that $\mu_A > 0$ and $\mu_B > 0$, so that it becomes $\mu_B + \mu_A(1 - \mu_B \rho_B \sigma_d^2) \geq 0$, which is satisfied owing to Lemma A.3 and Claim 3.

The argument for condition (57) is analogous, establishing that Assumption II.A3 holds.

It is easy to verify that the stepsize intervals provided in Fig. 5 match those from Stepsize rule I when β_{P} and β_{D} are given by (55). Therefore, it only remains to verify that that Relaxation parameter rule II is equivalent to Relaxation parameter rule I. First, observe that the definition of η' from Fig. 6 is obtained by plugging in β'_{P} and β'_{D} from (55) into Fig. 2. Moreover, as a consequence of Claim 2, $\min\left\{\frac{1}{\gamma}\beta_{\mathrm{P}}, \frac{1}{\tau}\beta_{\mathrm{D}}\right\} \leq \min\left\{\frac{1}{\gamma}\rho_A, \frac{1}{\tau}\mu_B\right\} \leq \eta'$ and the following assertions hold.

(i) If $\max\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \leq 0$, then $\frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}} \leq \eta'$.

(ii) if $\beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 0$, then $\frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma) \leq \eta'$ for any $\sigma \in (0, \|L\|]$, since

$$\theta_{\gamma\tau}(\sigma) = \sqrt{\left(\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}}\right)^2 - \tfrac{1}{\gamma\tau}\beta_{\mathrm{P}}\beta_{\mathrm{D}}(1 - \gamma\tau\sigma^2)} \geq \left|\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}}\right|.$$

The claimed equivalence follows immediately from these two assertions, completing the proof. □

Data/code Availability Supplementary Python code is available on GitHub.

Declarations

Conflict of interest The authors have no conflict of interest to declare.

References

1. Attouch, H., Théra, M.: A general duality principle for the sum of two operators. J. Convex Anal. **3**, 1–24 (1996)
2. Auslender, A., Teboulle, M.: Asymptotic Cones and Functions in Optimization and Variational Inequalities. Springer Science & Business Media, Berlin (2006). https://doi.org/10.1007/b97594
3. Baksalary, J.K., Pukelsheim, F., Styan, G.P.: Some properties of matrix partial orderings. Linear Algebra Appl. **119**, 57–85 (1989). https://doi.org/10.1016/0024-3795(89)90069-4
4. Balakrishnan, V., Vandenberghe, L.: Semidefinite programming duality and linear time-invariant systems. IEEE Trans. Autom. Control **48**(1), 30–41 (2003). https://doi.org/10.1109/TAC.2002.806652
5. Banert, S., Upadhyaya, M., Giselsson, P.: The Chambolle–Pock method converges weakly with $\theta > 1/2$ and $\tau\sigma\|l\|^2 < 4/(1+2\theta)$ (2023). https://doi.org/10.48550/arXiv.2309.03998
6. Bauschke, H.H., Moursi, W.M., Wang, X.: Generalized monotone operators and their averaged resolvents. Math. Program. **189**(1–2), 55–74 (2021). https://doi.org/10.1007/s10107-020-01500-6
7. Bernstein, D.S.: Matrix Mathematics. Princeton University Press, Princeton (2009). https://doi.org/10.1515/9781400833344
8. Boţ, R.I., Nguyen, D.K.: The proximal alternating direction method of multipliers in the nonconvex setting: convergence analysis and rates. Math. Oper. Res. **45**(2), 682–712 (2020). https://doi.org/10.1287/moor.2019.1008
9. Böhm, A.: Solving nonconvex-nonconcave min-max problems exhibiting weak Minty solutions. Trans. Mach. Learn. Res. (2023). https://openreview.net/forum?id=Gp0pHyUyrb
10. Boyd, S., El Ghaoui, L., Feron, E., Balakrishnan, V.: Linear Matrix Inequalities in System and Control Theory. SIAM (1994). https://doi.org/10.1137/1.9781611970777
11. Bredies, K., Chenchene, E., Lorenz, D.A., Naldi, E.: Degenerate preconditioned proximal point algorithms. SIAM J. Optim. **32**, 2376–2401 (2022). https://doi.org/10.1137/21M1448112
12. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011). https://doi.org/10.1007/s10851-010-0251-1
13. Combettes, P.L., Pennanen, T.: Proximal methods for cohypomonotone operators. SIAM J. Control Optim. **43**(2), 731–742 (2004). https://doi.org/10.1137/S0363012903427336
14. Condat, L.: A primal–dual splitting method for convex optimization involving Lipschitzian, proximable and linear composite terms. J. Optim. Theory Appl. **158**(2), 460–479 (2013). https://doi.org/10.1007/s10957-012-0245-9
15. Diakonikolas, J., Daskalakis, C., Jordan, M.: Efficient methods for structured nonconvex-nonconcave min-max optimization. In: Banerjee, A., Fukumizu, K. (eds.) Proceedings of The 24th International Conference on Artificial Intelligence and Statistics, Proceedings of Machine Learning Research, vol. 130, pp. 2746–2754. PMLR, Cambridge (2021)
16. Eckstein, J.: The Lions–Mercier splitting algorithm and the alternating direction method are instances of the proximal point algorithm. Tech. rep., LIDS-P-1769 (1988)
17. Eckstein, J., Bertsekas, D.P.: On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Program. **55**, 293–318 (1992). https://doi.org/10.1007/BF01581204
18. Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3**(4), 1015–1046 (2010). https://doi.org/10.1137/09076934X

19. Evens, B., Latafat, P., Patrinos, P.: Convergence of the Chambolle–Pock algorithm in the absence of monotonicity (2023). https://doi.org/10.48550/arXiv.2312.06540
20. Evens, B., Pas, P., Latafat, P., Patrinos, P.: Convergence of the preconditioned proximal point method and Douglas–Rachford splitting in the absence of monotonicity. Math. Program. (2025). https://doi.org/10.1007/s10107-024-02182-0
21. Giannessi, F.: On Minty variational principle. In: New Trends in Mathematical Programming: Homage to Steven Vajda, pp. 93–99. Springer US, Boston (1998). https://doi.org/10.1007/978-1-4757-2878-1_8
22. Golub, G.H., Van Loan, C.F.: Matrix Computations. JHU press, Baltimore (2013). https://doi.org/10.1137/1.9781421407944
23. Gorbunov, E., Taylor, A., Horváth, S., Gidel, G.: Convergence of proximal point and extragradient-based methods beyond monotonicity: the case of negative comonotonicity. In: Krause, A., Brunskill, E., Cho, K., Engelhardt, B., Sabato, S., Scarlett, J. (eds.) Proceedings of the 40th International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 202, pp. 11614–11641. PMLR, Cambridge (2023)
24. Grosz, J.: A note on the general Hermitian solution to AXA*= B. Bull. Malays. Math. Sci. Soc. **21**(2), 66 (1998)
25. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective. SIAM J. Imaging Sci. **5**(1), 119–149 (2012). https://doi.org/10.1137/100814494
26. Helmersson, A.: Methods for robust gain scheduling. Ph.D. thesis, Linköping University Electronic Press (1995)
27. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (2012). https://doi.org/10.1017/CBO9780511810817
28. Iusem, A.N., Pennanen, T., Svaiter, B.F.: Inexact variants of the proximal point algorithm without monotonicity. SIAM J. Optim. **13**(4), 1080–1097 (2003). https://doi.org/10.1137/S1052623401399587
29. Khatri, C.G., Mitra, S.K.: Hermitian and nonnegative definite solutions of linear matrix equations. SIAM J. Appl. Math. **31**(4), 579–585 (1976). https://doi.org/10.1137/0131050
30. Konnov, I.V.: A class of combined iterative methods for solving variational inequalities. J. Optim. Theory Appl. **94**(3), 677–693 (1997). https://doi.org/10.1023/A:1022605117998
31. Latafat, P., Patrinos, P.: Asymmetric forward–backward–adjoint splitting for solving monotone inclusions involving three operators. Comput. Optim. Appl. **68**(1), 57–93 (2017). https://doi.org/10.1007/s10589-017-9909-6
32. Li, G., Pong, T.: Global convergence of splitting methods for nonconvex composite optimization. SIAM J. Optim. **25**, 2434–2460 (2015). https://doi.org/10.1137/140998135
33. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**(6), 964–979 (1979). https://doi.org/10.1137/0716071
34. Martinet, B.: Régularisation d'inéquations variationnelles par approximations successives. Revue française d'informatique et de recherche opérationnelle. Série rouge **4**(R3), 154–158 (1970). https://doi.org/10.1051/m2an/197004R301541
35. Meijer, T.J., Holicki, T., Eijnden, Svd, Scherer, C.W., Heemels, W.M.: The non-strict projection lemma. IEEE Trans. Autom. Control **69**(8), 5584–5590 (2024). https://doi.org/10.1109/TAC.2024.3371374
36. Minty, G.J.: Monotone (nonlinear) operators in Hilbert space. Duke Math. J. **29**(3), 341–346 (1962). https://doi.org/10.1215/S0012-7094-62-02933-2
37. Mitra, S.K., Bhimasankaram, P., Malik, S.B.: Matrix Partial Orders, Shorted Operators and Applications, vol. 10. World Scientific, Singapore (2010). https://doi.org/10.1142/7170
38. Möllenhoff, T., Strekalovskiy, E., Moeller, M., Cremers, D.: The primal-dual hybrid gradient method for semiconvex splittings. SIAM J. Imaging Sci. **8**(2), 827–857 (2015). https://doi.org/10.1137/140976601
39. O'Connor, D., Vandenberghe, L.: On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. Math. Progr. **179**(1–2), 85–108 (2020). https://doi.org/10.1007/s10107-018-1321-1
40. Otero, R.G., Iusem, A.: Regularity results for semimonotone operators. Comput. Appl. Math. **30**(1), 13 (2011)
41. Pennanen, T.: Local convergence of the proximal point algorithm and multiplier methods without monotonicity. Math. Oper. Res. **27**(1), 170–191 (2002). https://doi.org/10.1287/moor.27.1.170.331
42. Penrose, R.: A generalized inverse for matrices. In: Mathematical Proceedings of the Cambridge Philosophical Society, vol. 51, pp. 406–413. Cambridge University Press (1955). https://doi.org/10.1017/S0305004100030401

43. Pethick, T., Fercoq, O., Latafat, P., Patrinos, P., Cevher, V.: Solving stochastic weak minty variational inequalities without increasing batch size. In: The Eleventh International Conference on Learning Representations (2023)
44. Pethick, T., Latafat, P., Patrinos, P., Fercoq, O., Cevher, V.: Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In: International Conference on Learning Representations (2021)
45. Quan, J., Evens, B., Sepulchre, R., Patrinos, P.: Scaled relative graphs for nonmonotone operators with applications in circuit theory (2024). https://doi.org/10.48550/arXiv.2411.17419
46. Rockafellar, R.T.: Augmented Lagrangians and applications of the proximal point algorithm in convex programming. Math. Oper. Res. **1**(2), 97–116 (1976). https://doi.org/10.1287/moor.1.2.97
47. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM J. Control. Optim. **14**(5), 877–898 (1976). https://doi.org/10.1137/0314056
48. Rockafellar, R.T.: Progressive decoupling of linkages in optimization and variational inequalities with elicitable convexity or monotonicity. Set-Valued Var. Anal. **27**(4), 863–893 (2019). https://doi.org/10.1007/s11228-018-0496-1
49. Rockafellar, R.T., Sun, J.: Solving monotone stochastic variational inequalities and complementarity problems by progressive hedging. Math. Progr. **174**(1–2), 453–471 (2019). https://doi.org/10.1007/s10107-018-1251-y
50. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis. Springer Science & Business Media, Berlin (2009). https://doi.org/10.1007/978-3-642-02431-3
51. Shefi, R., Teboulle, M.: Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization. SIAM J. Optim. **24**(1), 269–297 (2014). https://doi.org/10.1137/130910774
52. Solodov, M.V., Svaiter, B.F.: A hybrid projection-proximal point algorithm. J. Convex Anal. **6**(1), 59–70 (1999)
53. Solodov, M.V., Tseng, P.: Modified projection-type methods for monotone variational inequalities. SIAM J. Control. Optim. **34**(5), 1814–1830 (1996). https://doi.org/10.1137/S0363012994268655
54. Svaiter, B.F.: On weak convergence of the Douglas–Rachford method. SIAM J. Control. Optim. **49**(1), 280–287 (2011). https://doi.org/10.1137/100788100
55. Tian, Y.: Equalities and inequalities for inertias of Hermitian matrices with applications. Linear Algebra Appl. **433**(1), 263–296 (2010). https://doi.org/10.1016/j.laa.2010.02.018
56. Tian, Y.: Analytical solutions to some optimization problems on ranks and inertias of matrix-valued functions subject to linear matrix inequalities (2013). https://arxiv.org/abs/1301.0986
57. Yan, M., Li, Y.: On the improved conditions for some primal-dual algorithms. J. Sci. Comput. **99**(3), 74 (2024). https://doi.org/10.1007/s10915-024-02537-x
58. Yonglin, C., Bingjun, Z.: On g-inverses and the nonsingularity of a bordered matrix. Linear Algebra Appl. **133**, 133–151 (1990). https://doi.org/10.1016/0024-3795(90)90245-8
59. Zhu, M., Chan, T.: An efficient primal-dual hybrid gradient algorithm for total variation image restoration. Ucla Cam Report **34**, 8–34 (2008)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.