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

Brecht Evens Puya Latafat Panagiotis Patrinos

Abstract

The Chambolle–Pock algorithm (CPA), also known as the primal-dual hybrid gradient method (PDHG), has surged in popularity in the last decade due to its success in solving convex/monotone structured problems. This work provides convergence results 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 for CPA, 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, sufficient convergence conditions are obtained when the individual operators belong to the recently introduced class of semimonotone operators [19]. Since this class of operators encompasses many traditional operator classes including (hypo)- and co(hypo)monotone operators, this analysis recovers and extends existing results for CPA. Several examples are provided for the aforementioned problem classes to demonstrate and establish tightness of the proposed stepsize ranges.

Keywords. convex and nonconvex optimization · monotone and nonmonotone variational inequalities · inclusion problems · Chambolle–Pock · primal-dual hybrid gradient · semimonotone operators
AMS subject classifications. 47H04 · 49J52 · 49J53 · 65K05 · 65K15 · 90C26.

1 Introduction

This paper considers composite inclusion problems of the form

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

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, inclusion (P-I) corresponds to the first-order optimality condition of

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \;\; g(x) + h(Lx), \tag{1.1}$$

where $A = \partial g$ and $B = \partial h$ represent 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 [57, 18, 25]). 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.

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

Department of Electrical Engineering (ESAT-STADIUS) – KU Leuven, Kasteelpark Arenberg 10, 3001 Leuven, Belgium.
E-mail: {brecht.evens,puya.latafat,panos.patrinos}@kuleuven.be.

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.

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 [18, 25, 14, 30, 11] in the monotone setting. Many other widely used numerical methods can also be interpreted as special cases of PPPA, see e.g. [16, 45, 48, 17, 14]. 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) preconditioned proximal point algorithm applied to (G-I) consists of the following fixed point iterations.

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

By selecting a particular form for the preconditioner P and the operator T, the Chambolle–Pock algorithm can be retrieved. Specifically, consider the so-called *primal-dual inclusion*

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

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 the primal-dual inclusion (PD-I), with preconditioner

$$P = \begin{bmatrix} \frac{1}{\gamma}\mathrm{I}_n & -L^\top \\ -L & \frac{1}{\tau}\mathrm{I}_m \end{bmatrix}. \tag{1.2}$$

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, 45, 46, 47] but also for positive semidefinite ones [30, 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, 30, 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 $\gamma\tau\|L\|^2 \leq 1$.

Recently, convergence of PPPA in the nonmonotone setting has been considered in [19] under the assumption that T admits a set of oblique weak Minty solutions, defined as follows.

Definition 1.1 (V-oblique weak Minty solutions [19])**.** *An operator $T : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to have V-oblique weak Minty solutions at (a nonempty set) $\mathcal{S}^\star \subseteq \operatorname{zer} T$ for some symmetric matrix $V \in \mathbb{R}^{n\times n}$ if*

$$\langle v, z - z^\star \rangle \geq \mathrm{q}_V(v), \qquad \textit{for all } z^\star \in \mathcal{S}^\star, (z, v) \in \operatorname{gph} T, \tag{1.3}$$

where the quadratic form $\mathrm{q}_V(v) := \langle v, Vv \rangle$. Whenever $V = \rho\mathrm{I}$ for some $\rho \in \mathbb{R}$, we use the notation of 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, (1.3) reduces to the classic Minty variational inequality (MVI) [36, 20], 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 [15, 44, 43, 9, 22], as well as the classic PPA method [41, 27, 13, 22].

Leveraging the results from [19] 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 $V = \operatorname{blkdiag}(V_{\mathrm{P}}, V_{\mathrm{D}})$, which depends on the fundamental subspaces of L (see (3.3) 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 [19, Sec. 3] are retrieved.

In contrast to the setting of DRS, where the convergence results follow in a straightforward manner from those of PPPA (see proof of [19, 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 through 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 3.4).

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 $(\tilde{x}, \tilde{y}) \in \operatorname{gph} A$ if*

$$\langle x - \tilde{x}, y - \tilde{y}\rangle \geq \mathrm{q}_M(x - \tilde{x}) + \mathrm{q}_R(y - \tilde{y}), \qquad \text{for all } (x, y) \in \operatorname{gph} A, \tag{1.4}$$

where $\mathrm{q}_X(\cdot) \coloneqq \langle \cdot, \cdot\rangle_X$ 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 $(\tilde{x}, \tilde{y}) \in \operatorname{gph} A$. 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 (1.4) reduces to

$$\langle x - \tilde{x}, y - \tilde{y}\rangle \geq \mu\|x - \tilde{x}\|^2 + \rho\|y - \tilde{y}\|^2, \qquad \text{for all } (x, y) \in \operatorname{gph} A. \tag{1.5}$$

The class of (μ, ρ)-semimonotone operators was introduced in [19, 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 [19, Rem. 4.2 & Fig. 4]).

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 5.5 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 1.3 could be replaced by $(\mu_A\|L\|^2, \rho_A)$-semimonotonicity, which is in general a much more restrictive assumption.

Theorem 1.3 (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 $(\mu_B \mathrm{I}_m, \rho_B LL^\top)$-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.*

(i) (either) $\mu_A = \mu_B = 0$ and $\rho_A = \rho_B = 0$ (monotone case).

(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 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 $\operatorname{zer} T_{\mathrm{PD}}$.*

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

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)- and co(hypo)monotonicity. Secondly, by considering the more general class of semimonotone operators, we obtain fundamentally new convergence results (see case 1.3*(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 5.8).

As first observed in [50], 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 [31, 8]. This approach leads to requirements for L such as full row rank assumption, and restrictions on its condition number (see [8, Ass. 1, Rem. 2(c)]). Recently, a Lagrangian-based method with switching mechanism was developed in [24] for a more general class of nonconvex optimization problems. Notably, when restricting to the linear composite setting of (P-I), their work is the first able to circumvent these rank assumptions. 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.

1.1 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 3.4 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 3.6 and 3.7.

As our second main contribution, convergence results are provided for the class of semimonotone operators [19, 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 5.5). 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 Section 4 (see also Theorem 5.1).

1.2 Organization

The paper is structured in the following manner. In Section 1.3, some notation and standard definitions are provided. Section 2 recalls the main convergence results from [19] for PPPA in the nonmonotone setting. In Section 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. In Section 3.1, two particular examples are provided which demonstrate tightness of our main convergence theorem. Section 4 discusses and introduces various calculus rules for the class of (M, R)-semimonotone operators. Leveraging these calculus rules, Section 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, Section 6 concludes the paper. For the sake of readability, several proofs and auxiliary results are deferred to the Appendix.

1.3 Notation

The set of natural numbers including zero is denoted by $\mathbb{N} := \{0, 1, \ldots\}$. The set of real and extended-real numbers are denoted by $\mathbb{R} := (-\infty, \infty)$ and $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$, while the positive and strictly positive reals are $\mathbb{R}_+ := [0, \infty)$ and $\mathbb{R}_{++} := (0, \infty)$. We 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 $0_{m\times 0}0_{0\times n} = 0_{m\times n}$. 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 $\operatorname{tr} P$.

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 function $\mathrm{q}_P(x) := \langle x, Px\rangle$. Let $\operatorname{diag}(\cdot)$ denote the diagonal matrix whose arguments constitute its diagonal elements. For arbitrary matrices A and B, we define the direct sum $A \oplus B = \operatorname{blkdiag}(A, B)$, where $\operatorname{blkdiag}(\cdot)$ 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 $\operatorname{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}$. We say that f is proper if $\operatorname{dom} f \neq \emptyset$ and that f is lower semicontinuous (lsc) if the epigraph $\operatorname{epi} f := \{(x, \alpha) \in \mathbb{R}^n \times \mathbb{R} \mid f(x) \leq \alpha\}$ is a closed subset of $\mathbb{R}^{n+1}$. We denote the limiting subdifferential of f by ∂f. We denote the normal cone of a set $E \subseteq \mathbb{R}^n$ by N_E and the projection onto E is denoted by $\Pi_E(x) := \operatorname{arg\,min}_{z\in E} \|z - x\|$. 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 $\operatorname{dom} A := \{x \in \mathbb{R}^n \mid Ax \neq \emptyset\}$, its graph by $\operatorname{gph} A := \{(x, y) \in \mathbb{R}^n \times \mathbb{R}^d \mid y \in Ax\}$, and the set of its zeros by $\operatorname{zer} A := \{x \in \mathbb{R}^n \mid 0 \in Ax\}$. The inverse of A is defined through its graph: $\operatorname{gph} A^{-1} := \{(y, x) \mid (x, y) \in \operatorname{gph} A\}$. The *resolvent* of A is defined by $J_A := (\mathrm{id} + A)^{-1}$. We say that A is *outer semicontinuous (osc)* at $\tilde{x} \in \operatorname{dom} A$ if

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

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

Definition 1.4 ((co)monotonicity)**.** *An operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is said to be μ-monotone for some $\mu \in \mathbb{R}$ if*

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

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

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

A is said to be maximally (co-)monotone if its graph is not strictly contained in the graph of another (co-)monotone operator. We say that A is monotone if it is 0-monotone.

Definition 1.5 (parallel sum of operators)**.** *The parallel sum between operators $A, B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is defined as $A \mathbin{\square} B := (A^{-1} + B^{-1})^{-1}$.*

Definition 1.6 ([37, Def. 9.2.1] parallel sum of matrices)**.** *Let $X, Y \in \mathbb{R}^{n\times n}$ denote two matrices. We say that X and Y are parallel summable if*

$$\mathcal{R}(X) \subseteq \mathcal{R}(X + Y) \quad \textit{and} \quad \mathcal{R}(X^\top) \subseteq \mathcal{R}((X + Y)^\top),$$

or equivalently $\mathcal{R}(Y) \subseteq \mathcal{R}(X + Y)$ and $\mathcal{R}(Y^\top) \subseteq \mathcal{R}((X + Y)^\top)$. For parallel summable matrices X and Y, their parallel sum is defined as [37, Cor. 9.2.5]

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

If both X and Y are nonsingular, then, $X \mathbin{\square} Y = (X^{-1} + Y^{-1})^{-1}$.

Definition 1.7 (parallel sum of extended-real numbers)**.** *Let $a, b \in \overline{\mathbb{R}}$. We say that a and b are parallel summable if either $a = b = 0$ or $a + b \neq 0$ and their parallel sum is defined as*

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

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

2 Preliminaries on the preconditioned proximal point method

Departing from the classical monotone setting of [46], convergence of relaxed PPPA was established in [19] 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 reiterating it here. In particular, their 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 $(P + T)^{-1}P$ has full domain.*

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

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

$$\bar{\eta} := 1 + \lambda_{\min}(U^\top VPU)^2 > 0, \tag{2.1}$$

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

In contrast to the convergence analysis techniques relying on firm nonexpansiveness of the resolvent mapping, the analysis of [19] relies on a projective interpretation of the preconditioned proximal point algorithm, which dates back to [52, 51, 29]. Most notably, it was demonstrated in [19, Lem. 2.2] 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 contain 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 ([19, Thm. 2.3] convergence of PPPA)**.** *Suppose that Assumption I holds, and consider a sequence $(z^k, \bar{z}^k)_{k\in\mathbb{N}}$ generated by PPPA starting from $z^0 \in \mathbb{R}^n$ with relaxation parameters $\lambda_k \in (0, 2\bar{\eta})$ such that $\liminf_{k\to\infty} \lambda_k(2\bar{\eta} - \lambda_k) > 0$, where $\bar{\eta}$ is defined as in (2.1). Then, either a point $\bar{z}^k \in \operatorname{zer} T$ is reached in a finite number of iterations or the following hold for the sequence $(z^k, \bar{z}^k)_{k\in\mathbb{N}}$.*

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

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

(iii) The shadow sequences $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$, $(\Pi_{\mathcal{R}(P)}\, \bar{z}^k)_{k\in\mathbb{N}}$ are bounded.

Moreover, if $(P + T)^{-1}P$ is (single-valued) continuous then,

(iv) The limit points of $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$ are in $\Pi_{\mathcal{R}(P)} \operatorname{zer} T$.

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

As mentioned in [19], Assumption I.A3 can be further relaxed by only requiring (1.3) to hold on $(z, v) \in \operatorname{gph} T \cap (\mathcal{R}((P + T)^{-1}P) \times \mathcal{R}(P))$ instead. Under this relaxed assumption, all results from Theorem 2.1 remain valid, as the proof of Theorem 2.1 only involves invoking (1.3) at points in this restricted set. This relaxation will prove to be relevant in Example 3.6.

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} [18, 25, 14, 30, 11]. Relying upon the abstract duality framework from [1], [2, Sec. 6.9], this

[2] As $U^\top VPU$ is similar to a symmetric matric, its eigenvalues are real [19, Eqn. (2.9)].

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* $y \in T_{\mathrm{D}}$

(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 (1.2). Then, each iteration corresponds to first finding a solution for $\bar{x}^k$ and $\bar{y}^k$ to the inclusions

$$\tfrac{1}{\gamma}x^k - L^\top y^k \in (\tfrac{1}{\gamma}\mathrm{id} + A)\bar{x}^k \qquad \text{and} \qquad \tfrac{1}{\tau}y^k - Lx^k \in -2L\bar{x}^k + (\tfrac{1}{\tau}\mathrm{id} + B^{-1})\bar{y}^k$$

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 3.2 (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* (1.2)*. 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

$$L = \begin{bmatrix} Y & Y' \end{bmatrix} \begin{bmatrix} \Sigma & \\ & 0 \end{bmatrix} \begin{bmatrix} X^\top \\ X'^\top \end{bmatrix}, \quad Y = \begin{bmatrix} Y_1 & \cdots & Y_d \end{bmatrix}, \quad X = \begin{bmatrix} X_1 & \cdots & X_d \end{bmatrix}, \tag{3.1}$$

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 [21, Sec. 2.5.2]

$$\Pi_{\mathcal{R}(L)} = YY^\top, \quad \Pi_{\mathcal{R}(L^\top)} = XX^\top, \quad \Pi_{\mathcal{N}(L)} = \mathrm{I}_n - XX^\top = X'X'^\top \quad \text{and} \quad \Pi_{\mathcal{N}(L^\top)} = \mathrm{I}_m - YY^\top = Y'Y'^\top. \tag{3.2}$$

These projections will play a central role in our upcoming analysis. In particular, 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*

$$V := V_{\mathrm{P}} \oplus V_{\mathrm{D}} = \left(\beta_{\mathrm{P}}\, \Pi_{\mathcal{R}(L^\top)} + \beta'_{\mathrm{P}}\, \Pi_{\mathcal{N}(L)}\right) \oplus \left(\beta_{\mathrm{D}}\, \Pi_{\mathcal{R}(L)} + \beta'_{\mathrm{D}}\, \Pi_{\mathcal{N}(L^\top)}\right) \in \mathbb{S}^{n+m}, \tag{3.3}$$

and the following conditions hold, where $\gamma_{\min}$ *and* $\gamma_{\max}$ *are defined as in* (3.7)*:*

(i) $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$ *and* $[\beta'_{\mathrm{P}}]_-[\beta'_{\mathrm{D}}]_- < \frac{1}{\|L\|^2}$, *(ii)* $[-\beta'_{\mathrm{P}}]_+ < \gamma_{\max}$ *and* $[-\beta'_{\mathrm{D}}]_+ < \frac{1}{\gamma_{\min}\|L\|^2}$.

Note that in Assumption II.A3, the matrix V consists of two blocks V_{P} and V_{D}, each involving projections onto the range and the kernel of L and $L^\top$. This imposed structure on V 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 translates Assumption II.A3 to the properties of the associated primal and dual inclusions. This lemma extends [19, Lem. 3.2], which considers the case $L = \mathrm{I}$.

Lemma 3.3 (oblique weak Minty for primal and dual operator)**.** *Suppose that Assumption II.A3 holds and let*

$$\mathcal{S}^\star_P := \{x^\star \mid \exists y^\star : (x^\star, y^\star) \in \mathcal{S}^\star\} \subseteq \operatorname{zer} T_{\mathrm{P}} \quad \text{and} \quad \mathcal{S}^\star_D := \{y^\star \mid \exists x^\star : (x^\star, y^\star) \in \mathcal{S}^\star\} \subseteq \operatorname{zer} T_{\mathrm{D}}.$$

Then, the primal operator T_{P} *has* V_{P}*-oblique weak Minty solutions at* $\mathcal{S}^\star_P$ *and the dual operator* T_{D} *has* V_{D}*-oblique weak Minty solutions at* $\mathcal{S}^\star_D$*.*

Proof. Note that $\{((x_A, y_B), (y_A + L^\top y_B, x_B - Lx_A)) \mid (x_A, y_A) \in \operatorname{gph} A, (x_B, y_B) \in \operatorname{gph} B\}$ is equal to $\operatorname{gph} T_{\mathrm{PD}}$. Consequently, by Assumption II.A3 it holds 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

$$\left\langle y_A + L^\top y_B, x_A - x^\star \right\rangle + \left\langle x_B - Lx_A, y_B - y^\star \right\rangle \geq \mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B) + \mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A), \tag{3.4}$$

♠ In (3.4), 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

$$\left\langle y_A + L^\top y_B, x_A - x^\star \right\rangle \geq \mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B). \tag{3.5}$$

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 $y_B \in \operatorname{dom}(A^{-1} \circ (-L^\top)) \cap \operatorname{dom}(B^{-1}) = \operatorname{dom} T_{\mathrm{D}} \neq \emptyset$ and let $y_A = -L^\top y_B$ in (3.4). 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

$$\left\langle x_B - Lx_A, y_B - y^\star \right\rangle \geq \mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A). \tag{3.6}$$

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 □

Consequently, Lemma 3.3 implies that the blocks V_{P} and V_{D} from (3.3) can be interpreted as the primal and the dual blocks of V, respectively. As shown in the proof of Lemma 3.3, the quadratic terms $\mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B)$ and $\mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A)$ emerging in oblique weak Minty inequality correspond to the primal and the dual problems, respectively. By selecting V_{P} and V_{D} as in (3.3), these terms can be written as

$$\begin{aligned} \mathrm{q}_{V_{\mathrm{P}}}(y_A + L^\top y_B) &= \beta_{\mathrm{P}}\|\Pi_{\mathcal{R}(L^\top)}\, y_A + L^\top y_B\|^2 + \beta'_{\mathrm{P}}\|\Pi_{\mathcal{N}(L)}\, y_A\|^2, \\ \mathrm{q}_{V_{\mathrm{D}}}(x_B - Lx_A) &= \beta_{\mathrm{D}}\|\Pi_{\mathcal{R}(L)}\, x_B - Lx_A\|^2 + \beta'_{\mathrm{D}}\|\Pi_{\mathcal{N}(L^\top)}\, x_B\|^2, \end{aligned}$$

reducing to the norm of the scaled sum of a vector belonging to the range of $L^\top$ and another to its nullspace (resp., range of L and nullspace of $L^\top$). This decomposition proves essential in the proof of Theorem 3.4, as it enables to split condition (2.1) into two terms, one depending only on β_{P} and β_{D} and the other only depending on β'_{P} and β'_{D} (see (3.13)).

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 (1.2). To this end, the stepsizes γ and τ and the relaxation parameter λ need to adhere to certain conditions as well. These conditions are summarized in the following rules.

Table 1: *Definition of η' in Relaxation parameter rule I.*

	rank $L = n$	rank $L < n$
rank $L = m$	$+\infty$	$1 + \frac{1}{\gamma}\beta'_{\mathrm{P}}$
rank $L < m$	$1 + \frac{1}{\tau}\beta'_{\mathrm{D}}$	$\min\left\{1 + \frac{1}{\gamma}\beta'_{\mathrm{P}}, 1 + \frac{1}{\tau}\beta'_{\mathrm{D}}\right\}$

Stepsize rule I. *Define*

$$\delta := 1 + [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_{-}\left(\|L\|^2 - \sigma_d^2\right).$$

The stepsizes satisfy $\gamma \in (\max\{\gamma_{\min}, [-\beta'_{\mathrm{P}}]_{+}\}, \min\{\gamma_{\max}, \frac{1}{[-\beta'_{\mathrm{D}}]_{+}\|L\|^2}\})$ and $\tau \in (\max\{\tau_{\min}(\gamma), [-\beta'_{\mathrm{D}}]_{+}\}, \frac{1}{\gamma\|L\|^2}]$, where

$$\gamma_{\min} := \frac{2[-\beta_{\mathrm{P}}]_{+}}{\delta + \sqrt{\delta^2 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}}, \quad \gamma_{\max} := \frac{\delta + \sqrt{\delta^2 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2}}{2[-\beta_{\mathrm{D}}]_{+}\|L\|^2}, \quad \tau_{\min}(\gamma) := \frac{[-\beta_{\mathrm{D}}]_{+}(\gamma + \beta_{\mathrm{P}})}{\gamma(\delta - \beta_{\mathrm{P}}\beta_{\mathrm{D}}\|L\|^2) + \beta_{\mathrm{P}}}. \tag{3.7}$$

In the monotone setting corresponding to $\beta_{\mathrm{P}} = \beta_{\mathrm{D}} = \beta'_{\mathrm{P}} = \beta'_{\mathrm{D}} = 0$, this stepsize rule reduces to the classical stepsize rule $\gamma \in (0, +\infty)$ and $\tau \in (0, 1/\gamma\|L\|^2]$. On the other hand, when $L = \mathrm{I}$, it reduces to $\gamma \in (\gamma_{\min}, \gamma_{\max})$ and $\tau \in (\tau_{\min}(\gamma), 1/\gamma]$, where

$$\gamma_{\min} = \frac{2[-\beta_{\mathrm{P}}]_{+}}{1 + \sqrt{1 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}}}, \quad \gamma_{\max} := \frac{1 + \sqrt{1 - 4\beta_{\mathrm{P}}\beta_{\mathrm{D}}}}{2[-\beta_{\mathrm{D}}]_{+}}, \quad \tau_{\min}(\gamma) := \frac{[-\beta_{\mathrm{D}}]_{+}(\gamma + \beta_{\mathrm{P}})}{\gamma(1 - \beta_{\mathrm{P}}\beta_{\mathrm{D}}) + \beta_{\mathrm{P}}},$$

which matches the stepsize range for Douglas–Rachford splitting from [19, Thm. 3.3] (taking $\tau = 1/\gamma$).

Relaxation parameter rule I. *Define*

$$\theta_{\gamma\tau}(\sigma) := \sqrt{\left(\tfrac{1}{2\gamma}\beta_{\mathrm{P}} - \tfrac{1}{2\tau}\beta_{\mathrm{D}}\right)^2 + \beta_{\mathrm{P}}\beta_{\mathrm{D}}\sigma^2}. \tag{3.8}$$

Let $\bar\eta := \min\{\eta, \eta'\}$, where η' is defined as in Table 1 and

$$\eta := \begin{cases} \left.\begin{cases} 1 + \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0 \\ 1 + \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\|L\|), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} \geq 0 \end{cases}\right\} & \text{if } \gamma\tau < \frac{1}{\|L\|^2}, \\ 1 + \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d = 1, \\ \left.\begin{cases} 1 + \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_d), & \text{if } \beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0 \\ 1 + \frac{1}{2\gamma}\beta_{\mathrm{P}} + \frac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_2), & \text{if } \min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} \geq 0 \\ 1 + \frac{1}{\gamma}\beta_{\mathrm{P}} + \frac{1}{\tau}\beta_{\mathrm{D}}, & \text{otherwise} \end{cases}\right\} & \text{if } \gamma\tau = \frac{1}{\|L\|^2} \text{ and } d > 1. \end{cases} \tag{3.9}$$

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

Observe that in Relaxation parameter rule I there is an interplay between the stepsizes γ and τ and the range of admissible relaxation parameters λ. In the monotone setting $\beta_{\mathrm{P}} = \beta'_{\mathrm{P}} = \beta_{\mathrm{D}} = \beta'_{\mathrm{D}} = 0$ and this interplay vanishes as this relaxation rule reduces to the classical condition $\lambda \in (0, 2)$. 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 will be larger than two (see [19, 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 phenomenon will occur for small enough τ. Finally, when all β parameters are strictly positive then the upper bound for λ is larger than two for all valid stepsizes γ and τ (see e.g. Example 3.7).

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 (3.14)).

Theorem 3.4. *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 $\bar{z}^k \in \operatorname{zer} T_{\text{PD}}$ 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) $\bar{v}^k := P(z^k - \bar{z}^k) \in T_{\text{PD}}\bar{z}^k$ 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 $\operatorname{zer} T_{\text{PD}}$.

(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.

Moreover, if $J_{\gamma A}$ and $J_{\tau B^{-1}}$ are continuous, then

(iv) The limit points of $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$ are in $\Pi_{\mathcal{R}(P)} \operatorname{zer} T_{\text{PD}}$.

(v) If in Assumption II.A3, $\mathcal{S}^\star = \operatorname{zer} T_{\text{PD}}$, then $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$ converges to some element of $\Pi_{\mathcal{R}(P)} \operatorname{zer} T_{\text{PD}}$ and $(\bar{z}^k)_{k\in\mathbb{N}}$ converges to some element of $\operatorname{zer} T_{\text{PD}}$. Finally, if λ_k is uniformly bounded in the interval $(0,2)$, then $(z^k)_{k\in\mathbb{N}}$ converges to some element of $\operatorname{zer} T_{\text{PD}}$.

Proof. First, outer semicontinuity of T_{PD} follows from that of A and B [49, Theorem 5.7(a)], showing Assumption I.A1. Second, Assumption I.A2 holds since $J_{\gamma A}$ and $J_{\tau B^{-1}}$ having full domain implies that the preconditioned resolvent $(P + T_{\text{PD}})^{-1}P$ has full domain, owing to [49, Lemma 12.14]. Third, Assumption I.A3 is immediate. It only remains to show that Assumption I.A4 holds, i.e., that (2.1) holds. 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 defined as in (3.1). Let

$$Z := \begin{bmatrix} Z_1 & \cdots & Z_d \end{bmatrix} \in \mathbb{R}^{(m+n)\times(2r)},$$

which by construction has orthonormal columns. The preconditioner P can be decomposed as

$$P = \begin{bmatrix} \frac{1}{\gamma}\mathrm{I}_n & -L^\top \\ -L & \frac{1}{\tau}\mathrm{I}_m \end{bmatrix} \overset{(3.1)}{=} \begin{bmatrix} \frac{1}{\gamma}\mathrm{I}_n & -X\Sigma Y^\top \\ -Y\Sigma X^\top & \frac{1}{\tau}\mathrm{I}_m \end{bmatrix} \overset{(3.2)}{=} \underbrace{\begin{bmatrix} \frac{1}{\gamma}XX^\top & -X\Sigma Y^\top \\ -Y\Sigma X^\top & \frac{1}{\tau}YY^\top \end{bmatrix}}_{=: Z\hat{P}Z^\top} + \underbrace{\left(\tfrac{1}{\gamma}X'X'^\top\right) \oplus \left(\tfrac{1}{\tau}Y'Y'^\top\right)}_{=: Z'P'Z'^\top}, \tag{3.10}$$

where

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

By construction, Z' is an orthonormal basis for the range of $Z'P'Z'^\top$, since $P' \succ 0$. Let $\hat{U} = (\hat{U}_1 \otimes \mathrm{I}_{m_1}) \oplus \cdots \oplus (\hat{U}_d \otimes \mathrm{I}_{m_d})$, where $\hat{U}_i$ is an orthonormal basis for $\mathcal{R}(\hat{P}_i)$, for $i \in [d]$. Since $\hat{P}_i \succ 0$ for $i \in [2,d]$, select $\hat{U}_i = \mathrm{I}_2$ for $i \in [2,d]$, so that $\hat{U} = (\hat{U}_1 \otimes \mathrm{I}_{m_1}) \oplus \mathrm{I}_{(2r-2m_1)}$. Since $\hat{P}$ conforms to the same block-diagonal structure, $\hat{U}$ is an orthonormal basis for $\mathcal{R}(\hat{P})$. Moreover, since the columns of Z are orthonormal, $Z\hat{U}$ is an orthonormal basis for $\mathcal{R}(Z\hat{P}Z^\top)$. Consequently, an orthonormal basis for $\mathcal{R}(P)$ is given by

$$U = \begin{bmatrix} Z\hat{U} & Z' \end{bmatrix} = \begin{bmatrix} Z_1(\hat{U}_1 \otimes \mathrm{I}_{m_1}) & Z_2 & \cdots & Z_d & Z' \end{bmatrix}. \tag{3.11}$$

Analogous to (3.10), the V-oblique weak Minty matrix as defined in (3.3) can be decomposed as

$$V \overset{(3.2)}{=} \underbrace{\left(\beta_{\text{P}}XX^\top\right) \oplus \left(\beta_{\text{D}}YY^\top\right)}_{=: Z\hat{V}Z^\top} + \underbrace{\left(\beta'_{\text{P}}X'X'^\top\right) \oplus \left(\beta'_{\text{D}}Y'Y'^\top\right)}_{=: Z'V'Z'^\top}, \tag{3.12}$$

where

$$\hat{V}_i = \operatorname{diag}(\beta_{\text{P}}, \beta_{\text{D}}), \quad \hat{V} = \left((\hat{V}_1 \otimes \mathrm{I}_{m_1}) \oplus \cdots \oplus (\hat{V}_d \otimes \mathrm{I}_{m_d})\right) \quad \text{and} \quad V' = \beta'_{\text{P}}\mathrm{I}_{n-r} \oplus \beta'_{\text{D}}\mathrm{I}_{m-r}.$$

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 (3.10)-(3.12) that

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

As a result, condition (2.1) is equivalent to

$$\bar{\eta} = 1 + \lambda_{\min}(U^\top VPU) = \min\Big(\overbrace{1 + \lambda_{\min}(\hat{U}^\top \hat{V}\hat{P}\hat{U})}^{=:\eta}, \overbrace{1 + \lambda_{\min}(Z'^\top V'P'Z')}^{=:\eta'}\Big) > 0. \tag{3.13}$$

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

$$\begin{aligned}
\eta &= 1 + \min\left\{\lambda_{\min}(\hat{U}_i^\top \hat{V}_i\hat{P}_i\hat{U}_i)\right\}_{i=1}^d \\
&= 1 + \min\left\{\lambda_{\min}(\hat{U}_1^\top \hat{V}_1\hat{P}_1\hat{U}_1), \left\{\lambda_{\min}(\hat{V}_i\hat{P}_i)\right\}_{i=2}^d\right\} \\
&= 1 + \min\left\{\lambda_{\min}(\hat{U}_1^\top \begin{bmatrix} \frac{1}{\gamma}\beta_{\mathrm{P}} & -\beta_{\mathrm{P}}\sigma_1 \\ -\beta_{\mathrm{D}}\sigma_1 & \frac{1}{\tau}\beta_{\mathrm{D}} \end{bmatrix}\hat{U}_1), \left\{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\right\}_{i=2}^d\right\},
\end{aligned} \tag{3.14}$$

where $\theta_{\gamma\tau}(\cdot)$ is defined as in (3.8). Conversely, by definition of Z', V' and P', it holds that

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

which matches the definition of η' provided in Table 1. In what follows, condition $\eta > 0$ is studied for $\gamma\tau\|L\|^2 < 1$ and $1 - \gamma\tau\|L\|^2 = 0$, respectively.

♠ $\gamma\tau\|L\|^2 < 1$: Then, $\hat{P}_1 \succ 0$ so that $\hat{U}_1 = \mathrm{I}_2$ and

$$\eta = 1 + \min\left\{\lambda_{\min}(\hat{V}_i\hat{P}_i)\right\}_{i=1}^d = 1 + \min\left\{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\right\}_{i=1}^d,$$

which matches the definition of η provided in (3.9).

♠ $\gamma\tau\|L\|^2 = 1$: Then, the matrix $\hat{P}_1$ has $\frac{1}{\gamma} + \frac{1}{\tau}$ and zero as eigenvalues, so that

$$\hat{U}_1 = \sqrt{\tfrac{\tau}{\gamma+\tau}}\begin{bmatrix} 1 \\ -\sqrt{\gamma/\tau} \end{bmatrix} \quad \text{and} \quad \hat{P}_1 = (\tfrac{1}{\gamma} + \tfrac{1}{\tau})\hat{U}_1\hat{U}_1^\top. \tag{3.15}$$

Therefore,

$$\begin{aligned}
\eta &= 1 + \min\left\{\lambda_{\min}(\tfrac{\tau}{\gamma+\tau}\begin{bmatrix} 1 & -\sqrt{\gamma/\tau} \end{bmatrix}\begin{bmatrix} \frac{1}{\gamma}\beta_{\mathrm{P}} & -\beta_{\mathrm{P}}\sigma_1 \\ -\beta_{\mathrm{D}}\sigma_1 & \frac{1}{\tau}\beta_{\mathrm{D}} \end{bmatrix}\begin{bmatrix} 1 \\ -\sqrt{\gamma/\tau} \end{bmatrix}), \left\{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\right\}_{i=2}^d\right\} \\
&= 1 + \min\left\{\tfrac{1}{\gamma}\beta_{\mathrm{P}} + \tfrac{1}{\tau}\beta_{\mathrm{D}}, \left\{\tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\sigma_i)\right\}_{i=2}^d\right\},
\end{aligned}$$

where we used that $\sqrt{\gamma\tau}\sigma_1 = 1$. This matches the definition of η provided in (3.9).

It remains to show that $\bar{\eta} := \min\{\eta, \eta'\} > 0$ if and only if γ and τ are chosen according to Stepsize rule I.

♠ $\eta > 0$: By analyzing the six different cases from (3.9), it follows from Lemmas A.1 and A.2 that the set of pairs $(\gamma, \tau) \in \mathbb{R}^2_{++}$ satisfying $\gamma\tau \in (0, 1/\|L\|^2]$ and $\eta > 0$ is given by

$$\left\{(\gamma, \tau) \in \mathbb{R}^2 \mid \gamma \in (\gamma_{\min}, \gamma_{\max}), \tau \in (\tau_{\min}(\gamma), \tfrac{1}{\gamma\|L\|^2}]\right\} \tag{3.16}$$

In particular, for the first and fourth case this follows from Lemma A.2*(ii)*, for the second case this follows from Lemmas A.2*(i)* and A.2*(iii)* since $[\beta_{\mathrm{P}}]_-[\beta_{\mathrm{D}}]_- < \frac{1}{4\|L\|^2}$ by Assumption II.A3*(i)* and for the fifth case this follows from Lemma A.2*(i)*. Finally, for the third and the sixth case this follows from Lemma A.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 $\delta = 1$.

♠ $\eta' > 0$: By algebraic manipulation, it follows that the set of pairs $(\gamma, \tau) \in \mathbb{R}^2_{++}$ satisfying $\gamma\tau \in (0, 1/\|L\|^2]$ and $\eta' > 0$ is given by

$$\left\{(\gamma, \tau) \in \mathbb{R}^2 \mid \gamma \in (-[\beta'_{\mathrm{P}}]_-, \tfrac{1}{-[\beta'_{\mathrm{D}}]_-\|L\|^2}), \tau \in (-[\beta'_{\mathrm{D}}]_-, \tfrac{1}{\gamma\|L\|^2}]\right\}. \tag{3.17}$$

This set is nonempty iff $[\beta'_{\mathrm{P}}]_-[\beta'_{\mathrm{D}}]_- < \frac{1}{\|L\|^2}$, which is ensured by Assumption II.A3*(i)*.

♠ $\bar{\eta} := \min\{\eta, \eta'\} > 0$: As a consequence of the previous two results, the set of pairs $(\gamma, \tau) \in \mathbb{R}^2_{++}$ satisfying $\gamma\tau \in (0, {}^1\!/\!_{\|L\|^2}]$ and $\bar{\eta} > 0$ is given by the intersection of (3.16) and (3.17), i.e., by Stepsize rule I. This intersection is nonempty if and only if $-[\beta'_{\mathrm{P}}]_- < \gamma_{\max}$ and $\frac{1}{-[\beta'_{\mathrm{D}}]_- \|L\|^2} > \gamma_{\min}$, which is ensured by Assumption II.A3*(ii)*.

Consequently, Assumption I holds, and owing to Lemma 3.2 all claims for CPA follow from Theorem 2.1. □

Note that by a telescoping argument, a rate of $O(\frac{1}{N})$ can be obtained for $\min_{k=0,1,\dots,N} \|\bar{v}^k\|^2$ when λ_k is uniformly bounded in the interval $(0, 2\bar{\eta})$ (see [19, thm. 2.3(iv)]).

Observe that Theorem 3.4 discusses not only the convergence of $(z^k)_{k\in\mathbb{N}}$, but also of its projection onto the range of the preconditioner $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$. In particular, 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 that $\gamma\tau < {}^1\!/\!_{\|L\|^2}$, 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 identical. 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 *shadow sequence* $(s^k)_{k\in\mathbb{N}} := (x^k - \gamma y^k)_{k\in\mathbb{N}}$ to the set $\{x^\star - \gamma y^\star \mid (x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}\}$ [19, Thm. 3.3]. This is why in classical results for DRS, typically convergence of the sequences $(\bar{z}^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ is established as opposed to convergence of $(z^k)_{k\in\mathbb{N}}$ [33, 53]. In the following proposition, this shadow sequence interpretation is generalized to arbitrary L matrices.

Proposition 3.5 (convergent sequences)**.** *Using the SVD of L from (3.1), define the function $\psi : \mathbb{R}^{m+n} \to \mathbb{R}^{m+n-m_1}$ as*

$$f(x, y) := \left(X_1^\top x - \sqrt{\tfrac{\gamma}{\tau}} Y_1^\top y,\ X_{2:}^\top x,\ Y_{2:}^\top y\right), \tag{3.18}$$

where $X_{2:} := \begin{bmatrix} X_2 & \cdots & X_d & X' \end{bmatrix}$ and $Y_{2:} := \begin{bmatrix} Y_2 & \cdots & Y_d & Y' \end{bmatrix}$[3]*. 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*

$$s^k := \psi(x^k, y^k) \quad \textit{and} \quad \mathcal{T} := \left\{\psi(x^\star, y^\star) \,\middle|\, (x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}\right\}, \tag{3.19}$$

Then, the following statements hold.

(i) 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 $\mathcal{T}$.

(ii) The sequence $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$ converges to $\Pi_{\mathcal{R}(P)} \operatorname{zer} T_{\mathrm{PD}}$ if and only if $(s^k)_{k\in\mathbb{N}}$ converges to $\mathcal{T}$.

Proof. First, by plugging in (3.15) into (3.11), observe that

$$U = \begin{bmatrix} \sqrt{\tfrac{\tau}{\gamma+\tau}} Z_1 \begin{bmatrix} \mathrm{I}_{m_1} \\ -\sqrt{\gamma/\tau}\,\mathrm{I}_{m_1} \end{bmatrix} & Z_2 & \cdots & Z_d & Z' \end{bmatrix} = \begin{bmatrix} \sqrt{\tfrac{\tau}{\gamma+\tau}} X_1 & X_2 & 0 & \cdots & X_d & 0 & X' & 0 \\ -\sqrt{\tfrac{\gamma}{\gamma+\tau}} Y_1 & 0 & Y_2 & \cdots & 0 & Y_d & 0 & Y' \end{bmatrix}. \tag{3.20}$$

Therefore, ψ corresponds to the linear mapping

$$\psi(x, y) = \left(\sqrt{\tfrac{(\gamma+\tau)}{\tau}} \mathrm{I}_{m_1} \oplus \mathrm{I}_{m+n-2m_1}\right) U^\top \begin{bmatrix} x \\ y \end{bmatrix}.$$

As a result, the claims follow from (3.19), using that $\Pi_{\mathcal{R}(P)} = UU^\top$ and that U has orthonormal columns. □

Notably, Proposition 3.5 along with Theorem 3.4 establishes the convergence of an $(m+n-m_1)$-dimensional sequence $(s^k)_{k\in\mathbb{N}}$ of CPA when $\gamma\tau = \frac{1}{\|L\|^2}$. Since $s^k = x^k - \gamma y^k$ when $L = \mathrm{I}$ and $\tau = {}^1\!/\!_{\gamma}$, it follows immediately that Theorem 3.4 matches the convergence results for DRS obtained in [19, 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 3.7.

Finally, it is worth noting that, analogous to the analysis performed in [19, Sec. 3.3], it is possible to establish linear convergence of CPA for piecewise polyhedral mappings.

[3] When L has orthogonal rows (resp. columns) with identical norm, then $X_{2:}$ (resp. $Y_{2:}$) are empty and the terms $X_{2:}^\top x$ (resp. $Y_{2:}^\top y$) vanish.

3.1 Examples

In this subsection, two examples of CPA will be considered, demonstrating some of the main attributes of our obtained convergence results from Theorem 3.4. Supplementary Python code verifying these results can be found on GitHub[4] and the proofs are deferred to Appendix B.

In the first example, the tightness of the bounds on the relaxation parameter λ from Theorem 3.4 will be 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 on the relaxation parameter λ can be obtained by ensuring stability. Note that 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 does indeed have an impact on the convergence of CPA applied to this splitting, as becomes apparent in (3.22).

Example 3.6 (saddle point problem)**.** Consider the problem of finding a zero of the following structured linear inclusion

$$0 \in T_{\mathrm{P}} x = \begin{bmatrix} b\ell^2 & a \\ -a & b\ell^2 \end{bmatrix} x = \overbrace{\begin{bmatrix} 0 & a \\ -a & 0 \end{bmatrix}}^{A} x + \overbrace{\begin{bmatrix} \ell & 0 & 0 \\ 0 & \ell & 0 \end{bmatrix}}^{L^\top} \overbrace{\begin{bmatrix} b & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{bmatrix}}^{B} \overbrace{\begin{bmatrix} \ell & 0 \\ 0 & \ell \\ 0 & 0 \end{bmatrix}}^{L} x, \tag{3.21}$$

where $a, b, \ell \in \mathbb{R} \setminus \{0\}$ and $c \in \mathbb{R}$. Note that any solution to the inclusion problem $0 \in T_{\mathrm{P}} x$ is a minimax solution of $f(x_1, x_2) := a x_1 x_2 + \frac{b\ell^2}{2}(x_1^2 - x_2^2)$ 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 (3.21) with $\tau = \frac{1}{\gamma \|L\|^2}$ and fixed relaxation parameter λ. Then, the following assertions hold.

(i) By examining the spectral radius of the algorithmic operator, it can be seen that the sequence $(\Pi_{\mathcal{R}(P)} z^k)_{k \in \mathbb{N}}$ converges iff $\lambda \in \left(0, \bar{\lambda}\right)$ and that $(z^k)_{k \in \mathbb{N}}$ converges iff $\lambda \in \left(0, \min\{2, \bar{\lambda}\}\right)$, where

$$\bar{\lambda} := \min\left\{2\left(1 + \tfrac{b\ell^2}{\gamma(a^2 + b^2\ell^4)} + \tfrac{ba^2\ell^2\gamma}{a^2 + b^2\ell^4}\right), 2\left(1 + \gamma c \ell^2\right)\right\}. \tag{3.22}$$

This upper bound is strictly positive iff $1 + \gamma c \ell^2 > 0$ and either $b > 0$ or

$$b < 0, \ a^2 \neq b^2\ell^4 \text{ and } \gamma \in \left(\min\{-\tfrac{1}{b\ell^2}, -\tfrac{b\ell^2}{a^2}\}, \max\{-\tfrac{1}{b\ell^2}, -\tfrac{b\ell^2}{a^2}\}\right).$$

(ii) Theorem 3.4*(v)* is tight in the sense that it matches the bounds on the relaxation parameter λ from (3.22), when in (1.3) the vector v is restricted to $\mathcal{R}(P)$ (see the remark below Assumption I).

(iii) 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. An example of this is when $a = 10$, $b = c = -\frac{1}{4}$ and $\ell = 2$. Owing to Example 3.6*(i)*, the sequence $(z^k)_{k \in \mathbb{N}}$ then converges if and only if

$$\gamma \in \left(\tfrac{1}{100}, 1\right) \quad \text{and} \quad \lambda \in \left(0, 2 - \tfrac{2}{101\gamma} - \tfrac{200\gamma}{101}\right).$$

□

The second example focusses on a particular instance of Theorem 3.4. Specifically, it considers the case 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, Theorem 3.4 states that the admissible range for the relaxation parameter λ depends on the second largest singular value of L (see Relaxation parameter rule I). Although the following simple example involves n separable inclusions, it proves sufficient to demonstrate that this result is not merely a consequence of our analysis, but that this is also observed in practice.

Example 3.7 (influence of singular values)**.** Let $n \in \{2, 3, \ldots\}$ and $L = \operatorname{diag}(1, \ell_2 \ldots \ell_n)$, where $|\ell_k| < 1$, $\forall k \in \{2, \ldots, n\}$. Let

$$A = \operatorname{diag}\left(1, 1 + \sqrt{1 - \ell_2^2}, \ldots, 1 + \sqrt{1 - \ell_n^2}\right) \quad \text{and} \quad B = \operatorname{diag}\left(1, \tfrac{1}{1 + \sqrt{1 - \ell_2^2}}, \ldots, \tfrac{1}{1 + \sqrt{1 - \ell_n^2}}\right).$$

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 fixed relaxation parameter λ, where s^k is defined as in (3.19). Then, the following assertions hold.

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

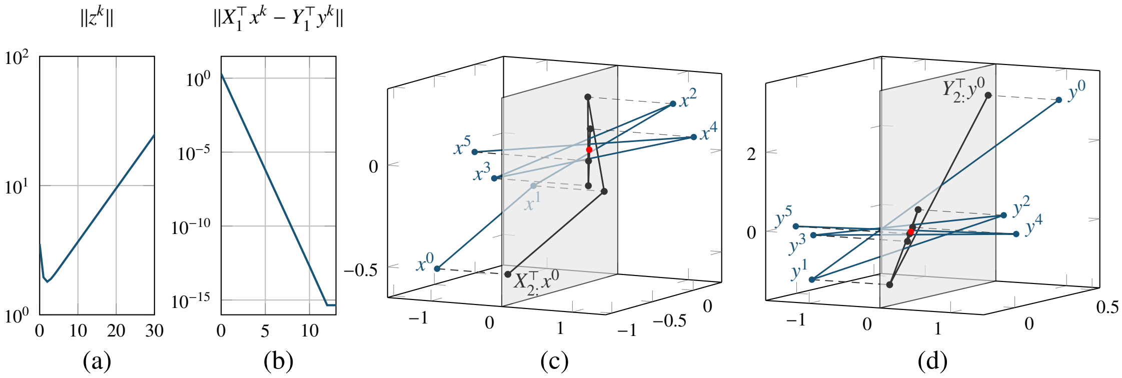

Figure 1: *Convergence of the sequence $(s^k)_{k\in\mathbb{N}} = (X_1^\top x^k - Y_1^\top y^k,\ X_{2:}^\top x^k,\ Y_{2:}^\top y^k)_{k\in\mathbb{N}}$ from Example 3.7 for $n = 3$, $\ell_2 = 1/2$, $\ell_3 = 1/5$ and $\lambda = 2.1$. (a) Norm of the sequence $(z^k)_{k\in\mathbb{N}} = (x^k, y^k)_{k\in\mathbb{N}}$. This sequence does not converge, since λ has been selected larger than two (see Theorem 3.4(v)). (b) Norm of the sequence $(X_1^\top x^k - Y_1^\top y^k)_{k\in\mathbb{N}}$, which converges to zero. (c) Visualization of the primal sequences $(x^k)_{k\in\mathbb{N}}$ and $(X_{2:}^\top x^k)_{k\in\mathbb{N}}$. It can be seen that 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 red dot). (d) 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.*

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

(ii) By Theorem 3.4*(v)* and Proposition 3.5, the sequences $(\Pi_{\mathcal{R}(P)}\, z^k)_{k\in\mathbb{N}}$ and $(s^k)_{k\in\mathbb{N}}$ converge to zero if λ is selected according to Relaxation parameter rule I, i.e., if $\lambda \in (0, \bar{\lambda})$, where

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

(iii) Let $n = 3$, $\ell_2 \in (0,1)$ and $\ell_3 = 1/5$. Then, by examining the spectral radius of the algorithmic operator, it can be seen that 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 3.4*(v)* (see Figure 2).

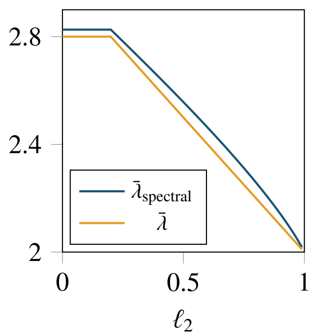

Figure 2: *The upper bounds $\bar{\lambda}$ and $\bar{\lambda}_{\text{spectral}}$ for Example 3.7, where $\bar{\lambda}_{\text{spectral}}$ is obtained by examining the spectral radius of the algorithmic operator.*

Finally, Figure 1 provides a numerical experiment demonstrating the convergence of the sequence $(s^k)_{k\in\mathbb{N}}$ for λ larger than two. □

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 [19, 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 Section 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.*

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

(i) The inverse operator A^{-1} is (maximally) (R_A, M_A)-semimonotone [at $(\tilde{y}, \tilde{x}) \in \operatorname{gph} A^{-1}$].

(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 $(\tilde{x}_A - u, w + \alpha\tilde{y}_A)$].

(iii) Operator $T := A \times B$ is (maximally) $(M_A \oplus M_B, R_A \oplus R_B)$-semimonotone [at $(\tilde{x}, \tilde{y}) \in \operatorname{gph} T$ where $\tilde{x} = (\tilde{x}_A, \tilde{x}_B)$ and $\tilde{y} = (\tilde{y}_A, \tilde{y}_B)$].

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 [19, Prop. 4.5] for (μ, ρ)-semimonotone operators.

Proposition 4.3 (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 [10, 26, 4, 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 [54, 55] and relies upon the classical results from [42, 28, 3, 23] for the linear matrix equality $D^\top X D = Y$.

Proposition 4.4 (symmetric solution of $D^\top X D \preceq Y$)**.** *Let $D \in \mathbb{R}^{m \times n}$ and $Y \in \mathbb{S}^n$. Then,*

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

$$\Pi_{\mathcal{N}(D)}\, Y\, \Pi_{\mathcal{N}(D)} \succeq 0 \quad \textit{and} \quad \operatorname{rank}(\Pi_{\mathcal{N}(D)}\, Y\, \Pi_{\mathcal{N}(D)}) = \operatorname{rank}(\Pi_{\mathcal{N}(D)}\, Y). \tag{4.1}$$

(ii) If (4.1) holds, then $X^\star \in C$, where

$$X^\star = \begin{bmatrix} 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} -Y & D^\top \\ D & 0 \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix} = (D^\dagger)^\top \Big(Y - Y\, \Pi_{\mathcal{N}(D)} (\Pi_{\mathcal{N}(D)}\, Y\, \Pi_{\mathcal{N}(D)})^\dagger\, \Pi_{\mathcal{N}(D)}\, Y \Big) D^\dagger. \tag{4.2}$$

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 4.3, the following corollary for linear operators is obtained.

Corollary 4.5 (linear operator)**.** *Let $D \in \mathbb{R}^{n \times n}$ and let $M \in \mathbb{S}^n$. Then,*

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

$$\Pi_{\mathcal{N}(D)}\, M\, \Pi_{\mathcal{N}(D)} \preceq 0 \quad \textit{and} \quad \operatorname{rank}(\Pi_{\mathcal{N}(D)}\, M\, \Pi_{\mathcal{N}(D)}) = \operatorname{rank}(\Pi_{\mathcal{N}(D)}(\tfrac{1}{2}D - M)). \tag{4.3}$$

(ii) If (4.3) holds, then D is $(M, R^\star)$-semimonotone, where

$$R^\star = \begin{bmatrix} 0 & \mathrm{I} \end{bmatrix} \begin{bmatrix} M - \frac{1}{2}(D + D^\top) & D^\top \\ D & 0 \end{bmatrix}^\dagger \begin{bmatrix} 0 \\ \mathrm{I} \end{bmatrix}. \tag{4.4}$$

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

$$R^\star = \tfrac{1}{2}(D + D^\top)^\dagger - {D^\dagger}^\top M D^\dagger + (D^\dagger)^\top M\, \Pi_{\mathcal{N}(D)} (\Pi_{\mathcal{N}(D)}\, M\, \Pi_{\mathcal{N}(D)})^\dagger\, \Pi_{\mathcal{N}(D)}\, M D^\dagger. \tag{4.5}$$

Note that $R^\star$ can be seen as the most optimal choice for R, as it solves the LMI from Proposition 4.3 as tightly as possible. A second consequence of Proposition 4.4 is the following result, which considers the semimonotonicity of an operator of the form $DTD^\top$.

Corollary 4.6 (semimonotonicity of $DTD^\top$)**.** *Let $D \in \mathbb{R}^{n\times m}$ and let operator $T : \mathbb{R}^m \rightrightarrows \mathbb{R}^m$ be (M, Y)-semimonotone [at $(D^\top\tilde{x}, \tilde{y}) \in \operatorname{gph} T$]. If (4.1) holds for D and T, then $DTD^\top$ is $(DMD^\top, X^\star)$-semimonotone [at $(\tilde{x}, D\tilde{y})$] where $X^\star$ is given by (4.2).*

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. First, consider the following set, which will be referred to as the effective domain of the parallel sum.

Definition 4.7 (effective domain of parallel sum)**.** *The set*

$$\operatorname{dom}_\square := \{(A, B) \in \mathbb{S}^n \times \mathbb{S}^n \mid A + B \succeq 0,\ A \text{ and } B \text{ are parallel summable}\} \tag{4.6}$$

is the effective domain of the parallel sum between two symmetric (possibly indefinite) matrices. Let $A = \alpha \mathrm{I}_n$ and $B = \beta \mathrm{I}_n$ where $\alpha, \beta \in \mathbb{R}$. Then, $(A, B) \in \operatorname{dom}_\square$ reduces to

$$(\alpha, \beta) \in \operatorname{dom}_\square = \{(\alpha, \beta) \mid \alpha + \beta > 0 \text{ or } \alpha = \beta = 0\}.$$

In the upcoming two propositions, 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. The first result generalizes [19, Prop. 4.7] for the sum of two (μ, ρ)-semimonotone operators.

Proposition 4.8 (sum and parallel sum)**.** *Let operator $A : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (M_A, R_A)-semimonotone [at $(\tilde{x}_A, \tilde{y}_A) \in \operatorname{gph} A$] and operator $B : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ be (M_B, R_B)-semimonotone [at $(\tilde{x}_B, \tilde{y}_B) \in \operatorname{gph} B$].*

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

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

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 4.9 (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 $(\tilde{x}, \tilde{y}) \in \operatorname{gph} T$], where $\mathcal{R}(R') \subseteq \mathcal{N}(D)$ and $(M, R) \in \operatorname{dom}_\square$. Then, $T + D$ is $(0, R' + M \square R)$-semimonotone [at $((\tilde{x}_A, \tilde{y}_B), (\tilde{y}_A + L^\top \tilde{y}_B, \tilde{x}_B - L\tilde{x}_A))$].*

5 Chambolle–Pock for semimonotone operators

In Section 3, convergence of CPA was established under an oblique weak Minty assumption on the underlying primal-dual operator. This section aims to provide 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) $(M_A, M_B) \in \operatorname{dom}_\square$, $(R_A, R_B) \in \operatorname{dom}_\square$, $\mathcal{R}(R'_A) \subseteq \mathcal{N}(L)$ and $\mathcal{R}(M'_B) \subseteq \mathcal{N}(L^\top)$.

(ii) *Operator A is $(L^\top M_A L, R_A + R'_A)$-semimonotone at $(x^\star, -L^\top y^\star) \in \operatorname{gph} A$.*

(iii) *Operator B is $(M_B + M'_B, L R_B L^\top)$-semimonotone at $(Lx^\star, y^\star) \in \operatorname{gph} B$.*

Then, T_{PD} has $((R'_A + R_A \mathbin{\square} R_B) \oplus (M'_B + M_A \mathbin{\square} M_B))$-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 4.2*(i)*, B^{-1} is $(L R_B L^\top, M_B + M'_B)$-semimonotone at $(y^\star, Lx^\star) \in \operatorname{gph} B^{-1}$, so that T is $(D^\top (M_A \oplus R_B) D, (R_A + R'_A) \oplus (M_B + M'_B))$-semimonotone at $((x^\star, y^\star), (-L^\top y^\star, Lx^\star)) \in \operatorname{gph} T$ due to Proposition 4.2*(iii)*. Consequently, by Lemma 4.9 it follows that T_{PD} is $(0, (R'_A + R_A \mathbin{\square} R_B) \oplus (M'_B + M_A \mathbin{\square} M_B))$-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. Then, by virtue of the particular form of $V = ((R'_A + R_A \mathbin{\square} R_B) \oplus (M'_B + M_A \mathbin{\square} M_B))$ from Theorem 5.1, the primal-dual operator T_{PD} has V-oblique weak Minty solutions at $\mathcal{S}^\star$, where V is given by (3.3) and

$$\beta_{\mathrm{P}} = \lambda_{\min}(X^\top (R_A \mathbin{\square} R_B) X), \quad \beta_{\mathrm{D}} = \lambda_{\min}(Y^\top (M_A \mathbin{\square} M_B) Y), \quad \beta'_{\mathrm{P}} = \lambda_{\min}(X'^\top R'_A X'), \quad \beta'_{\mathrm{D}} = \lambda_{\min}(Y'^\top M'_B Y'), \tag{5.1}$$

where X, X', Y, Y' are defined as in (3.1). Hence, Assumption II.A3 holds if the parameters from (5.1) satisfy II.A3*(i)* and II.A3*(ii)*.

5.1.1 Examples

As an implication of Theorem 5.1, consider the following result for the primal-dual operator emerging in constrained QP problems.

Example 5.2 (constrained QP)**.** Consider the following quadratic program

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \tfrac{1}{2} x^\top Q x + q^\top x \quad \text{subject to} \quad Lx \in C, \tag{5.2}$$

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 full column rank or $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$. Then, the following assertions hold.

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

(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 $\big(M_B(y^\star), 0\big)$-semimonotone at $(Lx^\star, y^\star) \in \operatorname{gph} B$.

(iii) If there exists a primal-dual pair $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ satisfying $(M_A, M_B(y^\star)) \in \operatorname{dom}_{\square}$, then Assumption II.A3 is satisfied for $\beta_{\mathrm{P}} = 0$, $\beta_{\mathrm{D}} = \lambda_{\min}(Y^\top (M_A \mathbin{\square} M_B(y^\star)) Y)$, $\beta'_{\mathrm{D}} = 0$ and $\beta'_{\mathrm{P}} = \lambda_{\min}\big(X'^\top Q^\dagger X'\big)$, provided that $[\beta_{\mathrm{D}}]_- [\beta'_{\mathrm{P}}]_- < \frac{1}{\|L\|^2}$.

Proof. See Appendix B. □

Leveraging this result, one can easily verify the underlying assumptions for Theorem 3.4, i.e., Assumption II, for (nonconvex) quadratic programs. Consider the following numerical example, where Theorem 3.4 is applied to a nonconvex QP with an indefinite Q matrix where L is full row rank. An example where L is rank-deficient is provided later in Example 5.7.

Example 5.3. Let $Q = \operatorname{diag}(1, -1, 2)$, $q = [\,-1\ 1\ -1\,]^\top$, $L = \begin{bmatrix} 1 & \frac{1}{4} & 0 \\ 0 & 1 & 0 \end{bmatrix}$ and $C := \big\{x \in \mathbb{R}^2 \mid 2 \le x_i \le 4, i = 1, 2\big\}$ in Example 5.2. Then, the global minimizer is given by $x^\star = [\,1\ 4\ \frac{1}{2}\,]^\top$ and the following assertions hold.

(i) Operator A is $(\operatorname{diag}(1, -1, 0), \operatorname{diag}(0, 0, \frac{1}{2}))$-semimonotone.

(ii) Operator B is $(\operatorname{diag}(0, \frac{3}{2}), 0)$-semimonotone at $(Lx^\star, -L^{\dagger\top} A x^\star) = \left(\begin{bmatrix} 2 \\ 4 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix}\right) \in \operatorname{gph} B$.

(iii) The primal-dual pair $(x^\star, -L^{\dagger\top} A x^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ is a V-oblique weak Minty solution of T_{PD} with V given by (3.3), where $\beta_{\mathrm{P}} = 0, \beta_{\mathrm{D}} = -3, \beta'_{\mathrm{P}} = \frac{1}{2}$ and $\beta'_{\mathrm{D}} = 0$ as in (5.1).

Table 2: *Range of the stepsizes γ and τ for CPA involving semimonotone operators.*

<table>
<tr><td colspan="2" rowspan="2"></td><td rowspan="2">$\mu_A = \mu_B = 0$</td><td colspan="2">$\mu_A + \mu_B > 0$</td></tr>
<tr><td>$\mu_A\mu_B \geq 0$</td><td>$\mu_A\mu_B < 0$</td></tr>
<tr><td colspan="2">$\rho_A = \rho_B = 0$</td><td colspan="2" rowspan="2">$\gamma \in (0, +\infty)$
$\tau \in (0, \frac{1}{\gamma\|L\|^2}]$</td><td>$\gamma \in (0, -\frac{1}{(\mu_A \square \mu_B)\|L\|^2})$
$\tau \in (-(\mu_A \square \mu_B), \frac{1}{\gamma\|L\|^2}]$</td></tr>
<tr><td rowspan="2">$\rho_A + \rho_B > 0$</td><td>$\rho_A\rho_B \geq 0$</td><td>$\gamma \in (0, \gamma_{\max})$
$\tau \in (\tau_{\min}(\gamma), \frac{1}{\gamma\|L\|^2}]$</td></tr>
<tr><td>$\rho_A\rho_B < 0$</td><td>$\gamma \in (-(\rho_A \square \rho_B), +\infty)$
$\tau \in (0, \frac{1}{\gamma\|L\|^2}]$</td><td>$\gamma \in (\gamma_{\min}, +\infty)$
$\tau \in (0, \frac{1}{\gamma\|L\|^2}]$</td><td>$\gamma \in (\gamma_{\min}, \gamma_{\max})$
$\tau \in (\tau_{\min}(\gamma), \frac{1}{\gamma\|L\|^2}]$</td></tr>
</table>

(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

$$\gamma \in \left(0, \tfrac{-1}{\beta_{\mathrm{D}}\|L\|^2}\right) \approx (0, 0.26), \quad \tau \in \left(-\beta_{\mathrm{D}}, \tfrac{1}{\gamma\|L\|^2}\right] \approx (3, \tfrac{0.779}{\gamma}], \quad \lambda \in (0, 2 + \tfrac{2}{\tau}\beta_{\mathrm{D}}) = (0, 2 - \tfrac{6}{\tau}),$$

where we used that $\|L\|^2 = \frac{33+\sqrt{65}}{32} \approx 1.28$. $\square$

Proof. The claimed assertions follow from those of Example 5.2 and Theorem 3.4, using that $L^\dagger = \left[\begin{smallmatrix} 1 & 0 & 0 \\ -\frac{1}{4} & 1 & 0 \end{smallmatrix}\right]^\top$, $\Pi_{\mathcal{R}(L^\top)} = \operatorname{diag}(1, 1, 0)$, $M_A = \frac{1}{16}\left[\begin{smallmatrix} 16 & -4 \\ -4 & -15 \end{smallmatrix}\right]$, $M_A \square \operatorname{diag}(0, \frac{3}{2}) = \operatorname{diag}(0, -3)$ and continuity of $J_{\gamma A}$ and $J_{\tau B^{-1}}$. $\square$

5.2 Sufficient conditions for convergence of CPA

Theorem 5.1 requires range conditions (5.1*(i)*) to hold for the semimonotonicity matrices of A and B. In this subsection, it is shown that this can be achieved by imposing a certain structure on the semimonotonicity 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* $(\mu_B \mathrm{I}_m, \rho_B LL^\top)$*-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.*

(i) (either) $\mu_A = \mu_B = 0$ *and* $\rho_A = \rho_B = 0$ *(monotone case).*

(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* $[\mu_A \square \mu_B]_-[\rho_A \square \rho_B]_- < \frac{1}{4\|L\|^2}$.

Owing to Theorem 5.1, Assumption III ensures that the primal-dual operator T_{PD} has oblique weak Minty solutions. This key result is stated in the following corollary.

Corollary 5.4. *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* (3.3) *and*

$$\beta_{\mathrm{P}} = \rho_A \square \rho_B, \quad \beta_{\mathrm{D}} = \mu_A \square \mu_B, \quad \beta_{\mathrm{P}}' = \begin{cases} 0, & \text{if } \operatorname{rank} L = n, \\ \rho_A, & \text{if } \operatorname{rank} L < n, \end{cases} \quad \beta_{\mathrm{D}}' = \begin{cases} 0, & \text{if } \operatorname{rank} L = m, \\ \mu_B, & \text{if } \operatorname{rank} L < m. \end{cases} \tag{5.3}$$

Proof. Observing that $(\mu_A, \mu_B) \in \operatorname{dom}_\square$ and $(\rho_A, \rho_B) \in \operatorname{dom}_\square$ and using that $\rho_A \mathrm{I}_n = \rho_A \Pi_{\mathcal{R}(L^\top)} + \beta_{\mathrm{P}}' \Pi_{\mathcal{N}(L)}$ and $\mu_B \mathrm{I}_m = \mu_B \Pi_{\mathcal{R}(L)} + \beta_{\mathrm{D}}' \Pi_{\mathcal{N}(L^\top)}$, the claim follows from Theorem 5.1. $\square$

Based on Corollary 5.4, one can thus easily verify Assumption II.A3 for CPA. Moreover, by plugging in the values for $\beta_{\mathrm{P}}, \beta_{\mathrm{P}}', \beta_{\mathrm{D}}, \beta_{\mathrm{D}}'$ from (5.3) into Stepsize rule I and Relaxation parameter rule I, the following simplified rules are obtained.

Table 3: *Definition of η' in Relaxation parameter rule II.*

	rank $L = n$	rank $L < n$
rank $L = m$	$+\infty$	$1 + \frac{1}{\gamma}\rho_A$
rank $L < m$	$1 + \frac{1}{\tau}\mu_B$	$\min\{1 + \frac{1}{\gamma}\rho_A, 1 + \frac{1}{\tau}\mu_B\}$

Stepsize rule II. *Let $\delta := 1 + [(\mu_A \square \mu_B)(\rho_A \square \rho_B)]_-(\|L\|^2 - \sigma_d^2)$. The stepsizes γ and τ satisfy the bounds provided in Table 2, where*

$$\gamma_{\min} := -\frac{2(\rho_A \square \rho_B)}{\delta + \sqrt{\delta^2 - 4(\mu_A \square \mu_B)(\rho_A \square \rho_B)\|L\|^2}}, \quad \gamma_{\max} := -\frac{\delta + \sqrt{\delta^2 - 4(\mu_A \square \mu_B)(\rho_A \square \rho_B)\|L\|^2}}{2(\mu_A \square \mu_B)\|L\|^2},$$

and $\tau_{\min}(\gamma) := -\dfrac{(\mu_A \square \mu_B)(\gamma + (\rho_A \square \rho_B))}{\gamma(\delta - (\mu_A \square \mu_B)(\rho_A \square \rho_B)\|L\|^2) + (\rho_A \square \rho_B)}$.

Relaxation parameter rule II. *Let η' be defined as in Table 3 and define*

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

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

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

Most notably, the resulting stepsize rule for CPA corresponds to a simple look-up table, analogous to the one from [19, Thm. 5.2] for DRS. The relaxation rule on the other hand is more involved, and depends on the choice of stepsizes γ and τ, the singular values of L and the semimonotonicity moduli of A and B.

Finally, based on the key result summarized in Corollary 5.4, the following corollary for the convergence of CPA for semimonotone operators is obtained.

Corollary 5.5 (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 3.4 hold.*

Proof. See Appendix B. □

Owing to the pointwise nature and parameter diversity of the underlying semimonotonicity assumptions, Corollary 5.5 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 5.6 (connection to existing theory)**.** Case (i) of Assumption III can be interpreted as a pointwise variant of the classical monotonicity assumption for CPA [12]. In case (ii) of Assumption III, a monotone problem is split in a nonmonotone fashion. In the optimization setting, this was already 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 5.5 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. Note that this case includes both results in the monotone and nonmonotone setting. □

5.2.1 Examples

Corollary 5.5 provides a set of sufficient conditions for the convergence of CPA which can be easily verified based on the calculus rules for semimonotone operators developed in Section 4 and [19, Sec. 4]. To demonstrate this, Corollary 5.5 will be applied to several examples previously discussed in this paper.

First of all, consider the following constrained QP, where Q is an indefinite matrix and L is rank-deficient. Previously, it was shown for the nonconvex QP from Example 5.3 that convergence of CPA can be established using Theorem 5.1 and Theorem 3.4. In this example, we show that if the monotonicity of A can be expressed in the form $\mu_A L^\top L$, then this result can be obtained directly using Corollary 5.5.

Example 5.7. Consider the QP from Example 5.2, where $Q = \operatorname{diag}(-3, -2, 1)$, $q = [\,0\ 1\ 0\,]^\top$, $L = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & -1 & 0 \end{bmatrix}$ and $C := \left\{x \in \mathbb{R}^2 \mid 1/2 \leq x_i \leq 1,\ i = 1, 2, 3\right\}$. Then, the global minimizer is given by $x^\star = [\,1\ 0\ 0\,]^\top$ and the following assertions hold.

(i) Operator A is $(\mu_A L^\top L, \operatorname{diag}(0, 0, 1))$-semimonotone where $\mu_A = -1$.

(ii) Operator B is $(\mu_B \mathrm{I}, 0)$-semimonotone at $(Lx^\star, -L^{\dagger\top} A x^\star) = ([\,1\ 1\ 1\,]^\top, [\,1\ 1\ 1\,]^\top) \in \operatorname{gph} B$where $\mu_B = 2$.

(iii) 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

$$\gamma \in (0, \tfrac{-1}{(\mu_A \,\square\, \mu_B)\|L\|^2}) = (0, \tfrac{1}{6}), \quad \tau \in (-(\mu_A \,\square\, \mu_B), \tfrac{1}{\gamma\|L\|^2}] = (2, \tfrac{1}{3\gamma}], \quad \lambda \in (0, 2 + \tfrac{2}{\tau}(\mu_A \,\square\, \mu_B)) = (0, 2 - \tfrac{4}{\tau}). \qquad \square$$

Proof. The claimed assertions follow from those of Example 5.2 and Corollary 5.5, using that A is $(\mu_A L^\top L, 0)$-semimonotone, $L^\dagger = \frac{1}{6}\begin{bmatrix} 2 & 2 & 2 \\ 0 & 3 & -3 \\ 0 & 0 & 0 \end{bmatrix}$ and continuity of $J_{\gamma A}$ and $J_{\tau B^{-1}}$. $\square$

Next, we revisit the two examples of Section 3.1, this time under the lens of semimonotonicity. First of all, consider the linear inclusion problem from Example 3.6, where the parameters a, b, c, l are selected as in Example 3.6*(iii)*.

Example 5.8 (saddle point problem (revisited))**.** Consider inclusion problem (3.21) with $a = 10$, $b = c = -\frac{1}{4}$ and $\ell = 2$. Using Proposition 4.3, 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 5.5, the sequence $(z^k)_{k\in\mathbb{N}}$ generated by applying CPA to (3.21) with $\tau = \frac{1}{\gamma\|L\|^2}$ and fixed relaxation parameter λ converges for

$$\gamma \in (\gamma_{\min}, \gamma_{\max}) \approx (0.055, 0.528) \quad \text{and} \quad \lambda \in \left(0, 2 - \tfrac{1}{10\gamma} - \tfrac{24\gamma}{7}\right). \qquad \square$$

The obtained range of stepsize parameters is only a subset of the tight range obtained in Example 3.6*(iii)*. 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 5.9 (influence of singular values (revisited))**.** Consider the composite inclusion problem $0 \in Ax + L^\top BLx$ from Example 3.7. It follows from Proposition 4.3 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 5.5 and Proposition 3.5, 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

$$\lambda \in \left(0, 2\left(1 + \tfrac{1}{2\gamma}\beta_{\mathrm{P}} + \tfrac{1}{2\tau}\beta_{\mathrm{D}} - \theta_{\gamma\tau}(\max\{|\ell_2|, \ldots, |\ell_n|\})\right)\right) = \left(0, 5/2 - 1/2 \max\{|\ell_2|, \ldots, |\ell_n|\}\right). \qquad \square$$

In Example 3.7*(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 5.4, 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 $A = 1/2 L^\top L + 1/2 A^\top A$ are lost in this process.

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 [19]) was introduced, and sufficient conditions for the convergence of CPA were provided for inclusion problems involving operators belonging to this class. Notably, 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 in the above-mentioned class of problems can be further extended 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 [32, 5], as well as analyzing other splitting methods in nonmonotone settings.

A Auxiliary lemmas

Lemma A.1 (solution of quadratic inequality)**.** *Let $\beta_{\mathrm{P}}, \beta_{\mathrm{D}} \in \mathbb{R}$, $\|L\| > 0, \sigma_d \in (0, \|L\|]$ and let $\delta := 1 + [\beta_{\mathrm{P}}\beta_{\mathrm{D}}]_-\big(\|L\|^2 - \sigma_d^2\big)$. Then, the following hold.*

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

(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 $\gamma \in (\gamma_{\min}, \gamma_{\max})$

Proof. See [19, Fact A.2]. □

Lemma A.2. *Let $\beta_{\mathrm{P}}, \beta_{\mathrm{D}} \in \mathbb{R}$, $\|L\| > 0$, $\sigma_d \in (0, \|L\|]$ and define the set*

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

where $\theta_{\gamma\tau}(\cdot)$ is defined as in (3.8). Then, the following hold.

(i) If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} > 0$ or $\beta_{\mathrm{P}}\beta_{\mathrm{D}} = 0$, then, for any $\sigma \in [\sigma_d, \|L\|]$, the set $\Gamma(\sigma)$ is nonempty and given by (3.16).

(ii) If $\beta_{\mathrm{P}}\beta_{\mathrm{D}} < 0$, then the set $\Gamma(\sigma_d)$ is nonempty and given by (3.16).

(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 $\Gamma(\|L\|)$ is given by (3.16).

Proof. 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

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

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 $\gamma\tau\Gamma_2(\sigma) = c_1(\sigma, \gamma)\tau + c_2(\gamma)$.

♠ Lemma A.2*(i)*: If $\min\{\beta_{\mathrm{P}}, \beta_{\mathrm{D}}\} > 0$, then $\Gamma_1 > 0$ and $\Gamma_2(\sigma) > 0$ since $1 - \gamma\tau\sigma^2 \geq 0$. On the other hand, if $\beta_{\mathrm{P}} = 0$ (resp. $\beta_{\mathrm{D}} = 0$), then $\Gamma_1 > 0$ and $\Gamma_2(\sigma) > 0$ iff $1 + \frac{1}{\gamma}\beta_{\mathrm{P}} > 0$ (resp. $1 + \frac{1}{\tau}\beta_{\mathrm{D}} > 0$). Therefore, it follows by algebraic manipulation that $\Gamma(\sigma)$ is nonempty and given by (3.16).

♠ Lemma A.2*(ii)*: If $\min\{\beta_{\mathrm{D}}, \beta_{\mathrm{P}}\} < 0$, 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 $1 - \gamma\tau\sigma_d^2 \geq 0$, so that it only remains to verify that $\Gamma_2(\sigma_d) > 0$.

◇ $\beta_{\mathrm{P}} > 0, \beta_{\mathrm{D}} < 0$: Then, $c_1(\sigma_d, \gamma) > 0$ and $c_2(\gamma) < 0$ and thus $(\gamma\tau)\Gamma_2(\sigma_d) > 0$ if and only if $\tau > {}^{-c_2(\gamma)}\!/_{c_1(\sigma_d,\gamma)} = \tau_{\min}(\gamma)$. The stepsize range for τ is nonempty if for some $\gamma > 0$ it holds that

$$\tau_{\min}(\gamma) < {}^1\!/_{\gamma\|L\|^2} \iff c_2(\gamma)\gamma\|L\|^2 + c_1(\sigma_d, \gamma) = \beta_{\mathrm{D}}\|L\|^2\gamma^2 + \Big(1 + \beta_{\mathrm{P}}\beta_{\mathrm{D}}(\|L\|^2 - \sigma_d^2)\Big)\gamma + \beta_{\mathrm{P}} > 0, \tag{A.1}$$

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

$$\Gamma(\sigma_d) = \Big\{(\gamma, \tau) \in \mathbb{R}^2 \;\Big|\; \gamma \in \big(0, \gamma_{\max}\big) \text{ and } \tau \in \Big(\tau_{\min}(\gamma), {}^1\!/_{\gamma\|L\|^2}\Big]\Big\} = (3.16).$$

$\diamond$ $\beta_P < 0$, $\beta_D > 0$: Observe that $\Gamma_2(\sigma_d) \leq 0$ for all $\gamma \in (0, {}^{-\beta_P}/_{(1-\beta_P\beta_D\sigma_d^2)}]$. Therefore, it holds that $\gamma > {}^{-\beta_P}/_{(1-\beta_P\beta_D\sigma_d^2)}$ and $c_1(\sigma_d, \gamma) > 0$, in which case $\Gamma_2(\sigma_d)$ is equivalent to $\tau > {}^{-c_2(\gamma)}/_{c_1(\sigma_d,\gamma)} = \tau_{\min}(\gamma)$. Ensuring that $\tau_{\min}(\gamma) < {}^{1}/_{\gamma\|L\|^2}$, i.e., solving (A.1) as before, yields $\gamma \in \left(\gamma_{\min}, +\infty\right)$. Finally, by observing that $\gamma_{\min} > -\beta_P > {}^{-\beta_P}/_{(1-\beta_P\beta_D\sigma_d^2)}$, it follows that $c_2(\gamma) \geq 0$ and thus $\tau_{\min}(\gamma) \leq 0$, so that

$$\Gamma(\sigma_d) = \left\{(\gamma, \tau) \in \mathbb{R}^2 \;\middle|\; \gamma \in \left(\gamma_{\min}, +\infty\right) \text{ and } \tau \in \left(0, {}^{1}/_{\gamma\|L\|^2}\right]\right\} = (3.16).$$

$\spadesuit$ Lemma A.2*(iii)*: First, observe that $\Gamma_1 > 0$ for $(\gamma, \tau) \in \mathbb{R}^2_{++}$ if and only if $\gamma > {}^{-\beta_P}/2$ and $\tau > {}^{-\gamma\beta_D}/_{(2\gamma+\beta_P)}$. As a result, the set $\Gamma(\|L\|)$ is empty when $\beta_P\beta_D \geq {}^{1}/_{\|L\|^2}$, as in this case the (quadratic) inequality ${}^{-\gamma\beta_D}/_{(2\gamma+\beta_P)} < {}^{1}/_{\gamma\|L\|^2}$ does not have a positive solution for $\gamma > {}^{-\beta_P}/2$. Consider the following cases, assuming that $\beta_P\beta_D < {}^{1}/_{\|L\|^2}$.

$\diamond$ $\gamma \in ({}^{-\beta_P}/2, {}^{-\beta_P}/_{(1-\beta_P\beta_D\|L\|^2)})$: Then, $\Gamma_2(\|L\|) > 0$ if and only if $\tau < \tau_{\min}(\gamma)$. Since it is easy to verify that in this case ${}^{-\gamma\beta_D}/_{(2\gamma+\beta_P)} > \tau_{\min}(\gamma)$, no such γ belong to the set $\Gamma(\|L\|)$.

$\diamond$ $\gamma = {}^{-\beta_P}/_{(1-\beta_P\beta_D\|L\|^2)}$: Then, $\Gamma_2(\|L\|) \leq 0$ for all $\tau > 0$.

$\diamond$ $\gamma > {}^{-\beta_P}/_{(1-\beta_P\beta_D\|L\|^2)}$: Then, $\Gamma_2(\|L\|) > 0$ if and only if $\tau > \tau_{\min}(\gamma) > {}^{-\gamma\beta_D}/_{(2\gamma+\beta_P)}$. The stepsize range for τ is nonempty if there exists some $\gamma > 0$ such that $\tau_{\min}(\gamma) < {}^{1}/_{\gamma\|L\|^2}$, which holds by Lemma A.1*(i)* if and only if $[\beta_P]_-[\beta_D]_- < {}^{1}/_{4\|L\|^2}$, in which case it follows from Lemma A.1*(ii)* that $\Gamma(\|L\|)$ is equal to (3.16). $\square$

Lemma A.3. *Let* $D := [\,\mathrm{I}_n \;\; \mathrm{I}_n\,]$, $(Y_1, Y_2) \in \operatorname{dom}_{\square}$ *and define* $Y = Y_1 \oplus Y_2$. *Then,* (4.1) *holds and* $X^\star$ *as defined in* (4.2) *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

$$\begin{aligned} \Pi_{\mathcal{N}(D)}\, Y &= (E \otimes \mathrm{I}_n)\begin{bmatrix} Y_1 & -Y_2 \end{bmatrix}, & \Pi_{\mathcal{N}(D)}\, Y\, \Pi_{\mathcal{N}(D)} &= (EE^\top) \otimes (Y_1 + Y_2), \\ \operatorname{rank}(\Pi_{\mathcal{N}(D)}\, Y) &= \operatorname{rank}\begin{bmatrix} Y_1 & Y_2 \end{bmatrix}, & \operatorname{rank}(\Pi_{\mathcal{N}(D)}\, Y\, \Pi_{\mathcal{N}(D)}) &= \operatorname{rank}(Y_1 + Y_2). \end{aligned}$$

Consequently, (4.1) 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 4.4*(ii)*, since

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

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. $\square$

Lemma A.4. *Suppose that Assumption III holds and that the sets* $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ *and* $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ *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 $\rho_A, \rho_B > 0$, for otherwise the two claims hold trivially. Consider $y_D \in \operatorname{dom}(A^{-1} \circ (-L^\top)) \cap \operatorname{dom}(B^{-1}) = \operatorname{dom} T_D \neq \emptyset$ 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

$$\langle x - x^\star, y + L^\top y^\star \rangle \geq \mathrm{q}_{\mu_A L^\top L}(x - x^\star) + \rho_A\|y + L^\top y^\star\|^2 \geq \mu_A\sigma_d^2\|\Pi_{\mathcal{R}(L^\top)}(x - x^\star)\|^2 + \rho_A\|y + L^\top y^\star\|^2, \quad \text{(A.2)}$$

where the involved norms are nonzero since $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ is not equal to the singleton $\{(-L^\top y^\star, \Pi_{\mathcal{R}(L^\top)} x^\star)\}$. 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

$$\langle x - x^\star, y + L^\top y^\star \rangle = \langle \Pi_{\mathcal{R}(L^\top)}(x - x^\star), y + L^\top y^\star \rangle \leq \mu_A\sigma_d^2\|\Pi_{\mathcal{R}(L^\top)}(x - x^\star)\|^2 + \tfrac{1}{4\mu_A\sigma_d^2}\|y + L^\top y^\star\|^2. \quad \text{(A.3)}$$

Combining (A.2) and (A.3), 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

$$\langle x - Lx^\star, y - y^\star \rangle \geq \mu_B \|x - Lx^\star\|^2 + \mathrm{q}_{\rho_B LL^\top}(y - y^\star) \geq \mu_B \|x - Lx^\star\|^2 + \rho_B \sigma_d^2 \|\Pi_{\mathcal{R}(L)}(y - y^\star)\|^2, \tag{A.4}$$

where the involved norms are nonzero since $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ 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

$$\langle x - Lx^\star, y - y^\star \rangle = \langle x - Lx^\star, \Pi_{\mathcal{R}(L)}(y - y^\star) \rangle \leq \mu_B \|x - Lx^\star\|^2 + \tfrac{1}{4\mu_B} \|\Pi_{\mathcal{R}(L)}(y - y^\star)\|^2. \tag{A.5}$$

Finally, combining (A.4) and (A.5), it follows that $\rho_B \sigma_d^2 \leq \frac{1}{4\mu_B}$, establishing the claim. □

Proposition A.5 (normal cone of a box)**.** *The normal cone operator $N_C : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ of the n-dimensional box $C := \{x \in \mathbb{R}^n \mid l_i \leq x_i \leq u_i,\ i = 1, \dots, n\}$ is $\left(\operatorname{diag}\left(\frac{|\tilde{v}_1|}{u_1 - l_1}, \dots, \frac{|\tilde{v}_n|}{u_n - l_n}\right), 0\right)$-semimonotone at $(\tilde{x}, \tilde{v}) \in \operatorname{gph} N_C$.*

Proof. By Proposition 4.2*(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

$$\tfrac{|\tilde{v}_i|}{u_i - \ell_i} |\tilde{x}_i - x_i|^2 \leq |\tilde{v}_i| |\tilde{x}_i - x_i| = \langle \tilde{v}_i, \tilde{x}_i - x_i \rangle.$$ □

B Omitted proofs

***Proof of Example 3.6** (saddle point problem).*

♠ 3.6*(i)*: By defining $H := \mathrm{I} + \lambda\Big((P + T_{\mathrm{PD}})^{-1} P - I\Big)$ and substituting $\tau = \frac{1}{\gamma \ell^2}$, 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 $\lambda \in \Big(0, \min\{2, \bar{\lambda}\}\Big)$. Analogously, the convergence result for $(\Pi_{\mathcal{R}(P)} z^k)_{k \in \mathbb{N}}$ can be obtained by analyzing the spectral radius of $\Pi_{\mathcal{R}(P)} H$.

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

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

owing to the Schur complement lemma. Therefore, when the vector v in (1.3) is restricted to $\mathcal{R}(P)$, Assumption II.A3 is equivalent to

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

where V is given by (3.3). 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 from (3.20) that

$$U = \begin{bmatrix} \frac{1}{\sqrt{1+\gamma^2\ell^2}} \mathrm{I}_2 & 0 \\ -\frac{\gamma\ell}{\sqrt{1+\gamma^2\ell^2}} \mathrm{I}_2 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad V = \beta_{\mathrm{P}} \mathrm{I}_2 \oplus \beta_{\mathrm{D}} \mathrm{I}_2 \oplus \beta_{\mathrm{D}}' \mathrm{I}_1,$$

where U is an orthonormal basis for $\mathcal{R}(P)$. As a result, (B.1) is satisfied if and only if

$$\begin{aligned} (T_{\mathrm{PD}}^{-1} U)^\top \Big(\tfrac{T_{\mathrm{PD}} + T_{\mathrm{PD}}^\top}{2} - T_{\mathrm{PD}}^\top V T_{\mathrm{PD}} \Big) T_{\mathrm{PD}}^{-1} U \succeq 0 \quad &\iff \quad \tfrac{b\ell^2(1+a^2\gamma^2) - (\beta_{\mathrm{P}} + \gamma^2\ell^2\beta_{\mathrm{D}})(a^2+b^2\ell^4)}{(a^2+b^2\ell^4)(1+\gamma^2\ell^2)} \mathrm{I}_2 \oplus (c - \beta_{\mathrm{D}}') \mathrm{I}_1 \succeq 0 \\ &\iff \quad \tfrac{1}{\gamma} \beta_{\mathrm{P}} + \gamma \ell^2 \beta_{\mathrm{D}} \leq \tfrac{b\ell^2(1+a^2\gamma)}{\gamma(a^2+b^2\ell^4)} \ \text{ and } \ \beta_{\mathrm{D}}' \leq c. \end{aligned}$$

Therefore, the upper bound on λ implied by Theorem 3.4 is given by

$$\lambda_{\max} := 2 \min\Big\{1 + \tfrac{1}{\gamma}\beta_{\mathrm{P}} + \gamma\ell^2\beta_{\mathrm{D}}, 1 + \gamma\ell^2\beta_{\mathrm{D}}'\Big\} \leq \min\Big\{2 + 2\tfrac{b\ell^2(1+a^2\gamma)}{\gamma(a^2+b^2\ell^4)}, 2(1 + \gamma c\ell^2)\Big\} = \bar{\lambda}.$$

♠ 3.6*(iii)*: For this particular instance $\operatorname{tr} T_{\mathrm{P}} = -2$ and $\operatorname{tr} T_{\mathrm{D}} = \operatorname{tr} T_{\mathrm{PD}} = -12$. Since the trace of a matrix equals the sum of its eigenvalues, the proof is completed. □

***Proof of* Example 3.7** *(influence of singular values).*

♠ 3.7*(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

$$\tfrac{T_{\mathrm{PD}}+T_{\mathrm{PD}}^\top}{2} - \tfrac{1}{2}T_{\mathrm{PD}}^\top T_{\mathrm{PD}} = \begin{bmatrix} A - \frac{1}{2}(A^\top A + L^\top L) & -\frac{1}{2}(AL - LB^{-1}) \\ -\frac{1}{2}(LA - B^{-1}L) & B^{-1} - \frac{1}{2}(B^{-1\top}B^{-1} + LL^\top) \end{bmatrix} \succeq 0. \tag{B.2}$$

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.

♠ 3.7*(ii)*: Follows from Theorem 3.4 and Proposition 3.5, using that $\|L\| = 1$ and $\gamma\tau = \frac{1}{\|L\|^2}$.

♠ 3.7*(iii)*: Analogous to the setting of Example 3.6*(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 $H := \mathrm{I} + \lambda\big((P + T_{\mathrm{PD}})^{-1}P - I\big)$ and P is defined as in (1.2). 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 $\lambda \in (0, \bar{\lambda}_{\mathrm{spectral}})$, where

$$\bar{\lambda}_{\mathrm{spectral}} \in \operatorname*{arg\,max}_{\lambda} \;\; \lambda \;\; \text{subject to} \;\; \|\Pi_{\mathcal{R}(P)} H(\lambda)\|_2 < 1.$$

The values for $\bar{\lambda}_{\mathrm{spectral}}$ reported in Figure 2 are obtained by solving this problem using SymPy. □

***Proof of* Proposition 4.1.** By the Fenchel-Young inequality, it holds for any $R \succ 0$ that

$$\langle x - \tilde{x}, y - \tilde{y}\rangle \geq \tfrac{1}{4}\,\mathrm{q}_{R^{-1}}(x - \tilde{x}) - \mathrm{q}_R(y - \tilde{y}), \qquad \text{for all } (x, y), (\tilde{x}, \tilde{y}) \in \operatorname{gph} A.$$

Therefore, (1.4) is always satisfied when $M \preceq \frac{1}{4}R^{-1}$. □

***Proof of* Proposition 4.2** *(inverting, shifting, scaling and cartesian product).*

♠ 4.2*(ii)*: First, consider the assertion where A is semimonotone only at $(\tilde{x}_A, \tilde{v}_A)$. Define $\tilde{s} = \tilde{x}_A - u$ and $\tilde{t} = y + \alpha\tilde{v}_A$, such that $(\tilde{s}, \tilde{t}) \in \operatorname{gph} T$. Then, it holds for all $(s, t) \in \operatorname{gph} T$ that

$$\begin{aligned} \langle s - \tilde{s}, t - \tilde{t}\rangle &= \alpha\Big\langle (s + u) - (\tilde{s} + u), \alpha^{-1}(t - y) - \alpha^{-1}(\tilde{t} - y)\Big\rangle \\ \text{(semimonotonicity of } A \text{ at } (\tilde{x}_A, \tilde{v}_A)) &\geq \alpha\,\mathrm{q}_M((s + u) - (\tilde{s} + u)) + \alpha\,\mathrm{q}_R(\alpha^{-1}(t - y) - \alpha^{-1}(\tilde{t} - y)) \\ &= \mathrm{q}_{\alpha M}(s - \tilde{s}) + \mathrm{q}_{\alpha^{-1}R}(t - \tilde{t}), \end{aligned}$$

where we used that $(\tilde{s} + u, \alpha^{-1}(\tilde{t} - y)) = (\tilde{x}_A, \tilde{v}_A)$ and $(s + u, \alpha^{-1}(t - y)) \in \operatorname{gph} A$. Hence, it follows that T is $(\alpha M_A, \alpha^{-1}R_A)$-semimonotone at $(\tilde{s}, \tilde{t})$.

If A is (M_A, R_A)-semimonotone at all $(\tilde{x}_A, \tilde{y}_A) \in \operatorname{gph} A$, we then know that T is $(\alpha M_A, \alpha^{-1}R_A)$-semimonotone at all points in the set $\{(\tilde{x}_A - u, y + \alpha\tilde{v}_A) \mid (\tilde{x}_A, \tilde{v}_A) \in \operatorname{gph} A\}$. Since this set is equal to $\operatorname{gph} T$, it follows that T is $(\alpha M_A, \alpha^{-1}R_A)$-semimonotone (everywhere).

♠ 4.2*(iii)*: Let A and B be semimonotone at respectively $(\tilde{x}_A, \tilde{v}_A)$ and $(\tilde{x}_B, \tilde{v}_B)$. Since $\operatorname{gph} T$ is equal to the set $\{((x_A, x_B), (v_A, v_B)) \mid v_A \in Ax_A, v_B \in Bx_B\}$, it holds for all $(x, v) \in \operatorname{gph} T$ that

$$\begin{aligned} \langle x - \tilde{x}, v - \tilde{v}\rangle &= \langle x_A - \tilde{x}_A, v_A - \tilde{v}_A\rangle + \langle x_B - \tilde{x}_B, v_B - \tilde{v}_B\rangle \\ \text{(semimonotonicity of } A \text{ and } B) &\geq \mathrm{q}_{M_A}(x_A - \tilde{x}_A) + \mathrm{q}_{R_A}(v_A - \tilde{v}_A) + \mathrm{q}_{M_B}(x_B - \tilde{x}_B) + \mathrm{q}_{R_B}(v_B - \tilde{v}_B) \\ &= \mathrm{q}_{M_A \oplus M_B}(x - \tilde{x}) + \mathrm{q}_{R_A \oplus R_B}(v - \tilde{v}). \end{aligned}$$

and thus T is $(M_A \oplus M_B, R_A \oplus R_B)$-semimonotone at $(\tilde{x}, \tilde{v}) \in \operatorname{gph} T$.

If A and B are semimonotone at all points in their graph, then T is $(M_A \oplus M_B, R_A \oplus R_B)$-semimonotone at all points in $\operatorname{gph} T$, which completes the proof. □

Proof of* Proposition 4.3 *(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 LMI in the statement. □

Proof of* Proposition 4.4 *(symmetric solution of $D^\top XD \preceq Y$).

♠ 4.4*(i)*: First, note that the problem of finding an $X \in \mathbb{S}^n$ such that $D^\top XD \preceq Y$ is equivalent to the problem of finding a pair $(X, Z) \in \mathbb{S}^n \times \mathbb{S}^m$ such that

$$D^\top XD = Y - Z \tag{B.3}$$

and $Z \succeq 0$. Second, observe that by [23, 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 [28, Thm. 2.2], a matrix $Z \succeq 0$ satisfying this condition exists if and only if (4.1) holds, and the general solution is given by

$$Z = Y \Pi_{\mathcal{N}(D)}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y + D^\top(D^\dagger)^\top G D^\top (D^\dagger)^\top, \tag{B.4}$$

where $G \in \mathbb{S}^n$ is an arbitrary symmetric positive semidefinite matrix.

♠ 4.4*(ii)*: Substituting (B.4) into (B.3) yields

$$D^\top XD = Y - Y \Pi_{\mathcal{N}(D)}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y - D^\top(D^\dagger)^\top G D^\top (D^\dagger)^\top,$$

of which the general solution is given by [23, Prop. 1]

$$X = \underbrace{(D^\dagger)^\top(Y - Y \Pi_{\mathcal{N}(D)}(\Pi_{\mathcal{N}(D)} Y \Pi_{\mathcal{N}(D)})^\dagger \Pi_{\mathcal{N}(D)} Y - G)D^\dagger}_{= X^\star - (D^\dagger)^\top G D^\dagger} + H - \Pi_{\mathcal{R}(D)} H \Pi_{\mathcal{R}(D)} \tag{B.5}$$

where $H \in \mathbb{S}^n$ is an arbitrary matrix. Substituting (B.5) into $D^\top XD \preceq Y$ shows that

$$Y - D^\top XD = Y - D^\top X^\star D + \Pi_{\mathcal{R}(D^\top)} G \Pi_{\mathcal{R}(D^\top)} \succeq Y - D^\top X^\star D \succeq 0.$$

Finally, the alternative expression for $X^\star$ given in (4.2) follows directly from [56, Lemma 3].

♠ 4.4*(iii)*: [7, Fact 6.4.38] □

Proof of* Corollary 4.6 *(semimonotonicity of $DTD^\top$). If (4.1) holds, then it follows from Proposition 4.4 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 $(\tilde{x}, D\tilde{y})$].

First, consider the case where T is semimonotone only at a single point $(D^\top\tilde{x}, \tilde{y})$. Let $(D^\top x, y) \in \operatorname{gph} T$ and denote $u = Dy$ and $\tilde{u} = D\tilde{y}$. Then, $(x, u), (\tilde{x}, \tilde{u}) \in \operatorname{gph} DTD^\top$ and it holds that

$$\begin{aligned}
\langle x - \tilde{x}, u - \tilde{u}\rangle &= \langle D^\top(x - \tilde{x}), y - \tilde{y}\rangle \\
\text{(semi. of } T \text{ at } (D^\top\tilde{x}, \tilde{y})) &\geq \mathrm{q}_M(D^\top(x - \tilde{x})) + \mathrm{q}_Y(y - \tilde{y}) \\
&\geq \mathrm{q}_{DMD^\top}(x - \tilde{x}) + \mathrm{q}_{X^\star}(D(y - \tilde{y})) = \mathrm{q}_{DMD^\top}(x - \tilde{x}) + \mathrm{q}_{X^\star}(u - \tilde{u}),
\end{aligned}$$

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 $(\tilde{x}, D\tilde{y})$.

Hence, if T is (M, Y)-semimonotone at all $(D^\top\tilde{x}, \tilde{y}) \in \operatorname{gph} T$, then $DTD^\top$ is $(DMD^\top, X^\star)$-semimonotone at all points in $\{(\tilde{x}, D\tilde{y}) \mid (D^\top\tilde{x}, \tilde{y}) \in \operatorname{gph} T\}$, which equals $\operatorname{gph} DTD^\top$. □

Proof of* Proposition 4.8 *(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 4.2*(iii)*, operator T is $(M, R) = (M_A \oplus M_B, R_A \oplus R_B)$-semimonotone [at $((\tilde{x}, \tilde{x}), (\tilde{y}_A, \tilde{y}_B)) \in \operatorname{gph} T$]. Consequently, it follows from Corollary 4.6 and Lemma A.3 that $DTD^\top = A + B$ is $(M_A + M_B, R_A \,\Box\, R_B)$-semimonotone [at $(\tilde{x}, \tilde{y}_A + \tilde{y}_B)$]. Finally, the claim for the parallel sum follows directly from those for the sum and Proposition 4.2*(i)*, since $A \,\Box\, B := (A^{-1} + B^{-1})^{-1}$. □

Proof of Lemma 4.9 *(sum with (skew-)symmetric matrix).* First, consider the assertion where T is semimonotone at $(\tilde{x}, \tilde{y})$. Let $(x, y) \in \operatorname{gph} T$. Then, $(x, y + Dx), (\tilde{x}, \tilde{y} + D\tilde{x}) \in \operatorname{gph}(T + D)$ and

$$\langle x - \tilde{x}, y - \tilde{y} + D(x - \tilde{x})\rangle \geq \mathfrak{q}_{D^\top MD}(x - \tilde{x}) + \mathfrak{q}_{R+R'}(y - \tilde{y}) + \langle x - \tilde{x}, D(x - \tilde{x})\rangle.$$

By Corollary 4.5*(ii)* and skew-symmetry of D, it follows that D is $(-D^\top MD, \Pi_{\mathcal{R}(D)} M \Pi_{\mathcal{R}(D)})$-semimonotone. Consequently,

$$\begin{aligned}\langle x - \tilde{x}, y - \tilde{y} + D(x - \tilde{x})\rangle &\geq \mathfrak{q}_{R+R'}(y - \tilde{y}) + \mathfrak{q}_{\Pi_{\mathcal{R}(D)} M \Pi_{\mathcal{R}(D)}}(Dx - D\tilde{x}) \\ &= \mathfrak{q}_{R'}(y - \tilde{y} + D(x - \tilde{x})) + \mathfrak{q}_R(y - \tilde{y}) + \mathfrak{q}_M(D(x - \tilde{x})) \\ &\geq \mathfrak{q}_{R'+M \square R}(y - \tilde{y} + D(x - \tilde{x}))\end{aligned}$$

where the fact that $\mathcal{R}(R') \subseteq \mathcal{N}(D)$ was used in the equality, and the final inequality follows from Proposition 4.4. Finally, when T is semimonotone, the claim follows analogously by considering all $(\tilde{x}, \tilde{y}) \in \operatorname{gph} T$, completing the proof. □

Proof of Example 5.2 *(constrained QP).*

♠ 5.2*(i)*: Let $M := L^\top M_A L = \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)}$ and observe that

$$\Pi_{\mathcal{N}(Q)} M = \Pi_{\mathcal{N}(Q)} \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)} = \Pi_{\mathcal{N}(Q)}(\mathrm{I}_n - \Pi_{\mathcal{N}(L)}) Q \Pi_{\mathcal{R}(L^\top)} = 0, \tag{B.6}$$

where the final equality holds since $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$. Therefore, (4.3) is satisfied for $D = Q$ and Q is $(L^\top M_A L, R^\star)$-semimonotone owing to Corollary 4.5*(ii)*, where $R^\star$ is given by (4.5), i.e. $R^\star = Q^\dagger - Q^\dagger M Q^\dagger = Q^\dagger - Q^\dagger \Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{R}(L^\top)} Q^\dagger$, where we used (B.6) and symmetry of Q. Moreover, since $\Pi_{\mathcal{R}(L^\top)} Q \Pi_{\mathcal{N}(L)} = 0$, it holds owing to [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 4.2*(ii)*.

♠ 5.2*(ii)*: Owing to 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.5.

♠ 5.2*(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, using that $\Pi_{\mathcal{N}(L)} = X' X'^\top$, it follows that the parameters $\beta_{\mathrm{P}}, \beta_{\mathrm{D}}, \beta'_{\mathrm{P}}$ and β'_{D} given in 5.2*(iii)* match those from (5.1). As remarked below Theorem 5.1, Assumption II.A3 holds if these parameters satisfy II.A3*(i)* and II.A3*(ii)*, i.e., if $[-\beta'_{\mathrm{P}}]_+ < \gamma_{\max} = \frac{1}{[-\beta_{\mathrm{D}}]_+ \|L\|^2}$, completing the proof. □

Proof of Corollary 5.5 *(convergence of* CPA *under semimonotonicity).* Note that Assumption III implies Assumption II.A3 by Corollary 5.4, where $\beta_{\mathrm{P}}, \beta'_{\mathrm{P}}, \beta_{\mathrm{D}}, \beta'_{\mathrm{D}} \in \mathbb{R}$ are given by (5.3). Therefore, it only remains to show that Stepsize rule II is equivalent to Stepsize rule I and that Relaxation parameter rule II is equivalent to Relaxation parameter rule I.

For technical reasons soon to be clear, we first show that $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ and $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ cannot be singletons. Suppose to the contrary that either $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ or $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ is a singleton. Since $\operatorname{zer} T_{\mathrm{PD}}$ is assumed to be nonempty, by Proposition 3.1 for all $(x^\star, y^\star) \in \operatorname{zer} T_{\mathrm{PD}}$ it holds that $(-L^\top y^\star, \Pi_{\mathcal{R}(L^\top)} x^\star) \in \operatorname{gph} \Pi_{\mathcal{R}(L^\top)} A^{-1}$ and $(Lx^\star, \Pi_{\mathcal{R}(L)} y^\star) \in \operatorname{gph} \Pi_{\mathcal{R}(L)} B$. Therefore, since either $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ or $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ is a singleton, it follows that $(\Pi_{\mathcal{R}(L^\top)} \oplus \Pi_{\mathcal{R}(L)}) \operatorname{zer} T_{\mathrm{PD}}$ is a singleton, which in turn implies that both $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ and $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ are singletons. Moreover, it holds that

$$\begin{aligned}\operatorname{gph}(J_{\gamma A}) &= \left\{(x + \gamma y, x) \mid (y, x) \in \operatorname{gph} A^{-1}\right\} \\ &= \left\{(x - \gamma L^\top y^\star, x) \mid x \in A^{-1}(-L^\top y^\star)\right\} \\ &= \left\{(\Pi_{\mathcal{R}(L^\top)} x^\star - \gamma L^\top y^\star + x', \Pi_{\mathcal{R}(L^\top)} x^\star + x') \mid x' \in \Pi_{\mathcal{N}(L)} A^{-1}(-L^\top y^\star)\right\},\end{aligned} \tag{B.7}$$

where we used that $\operatorname{dom}(A^{-1}) = \operatorname{dom}(\Pi_{\mathcal{R}(L^\top)} A^{-1}) = -L^\top y^\star$ in the second equality and that $\mathcal{R}(L^\top)$ and $\mathcal{N}(L)$ are orthogonal complements in the final equality. Using an analogous argument, it follows that

$$\operatorname{gph}(J_{\tau B^{-1}}) = \left\{(\Pi_{\mathcal{R}(L)} y^\star + \tau L x^\star + y', \Pi_{\mathcal{R}(L)} y^\star + y') \mid y' \in \Pi_{\mathcal{N}(L^\top)} B(Lx^\star)\right\}. \tag{B.8}$$

Consequently, the resolvents $J_{\gamma A}$ and $J_{\tau B^{-1}}$ do not have full domain, contradicting Assumption II.A2.

Having shown that the sets $\operatorname{gph}(\Pi_{\mathcal{R}(L^\top)} A^{-1})$ and $\operatorname{gph}(\Pi_{\mathcal{R}(L)} B)$ are not singletons, it follows from Lemma A.4 that $[\mu_A]_+[\rho_A]_+ \leq 1/4\sigma_d^2$ and $[\mu_B]_+[\rho_B]_+ \leq 1/4\sigma_d^2$. Additionally, it holds by Assumption III that $(\mu_A, \mu_B) \in \operatorname{dom}_\square$ and $(\rho_A, \rho_B) \in \operatorname{dom}_\square$, so that by definition

$$\beta_\mathrm{P} = \rho_A \square \rho_B \leq \rho_A \quad \text{and} \quad \beta_\mathrm{D} = \mu_A \square \mu_B \leq \mu_B. \tag{B.9}$$

As a consequence of Lemma A.4 and (B.9), we claim that under Assumption III Stepsize rule I for γ and τ reduces to $\gamma \in (\gamma_{\min}, \gamma_{\max})$ and $\tau \in (\tau_{\min}(\gamma), \frac{1}{\gamma\|L\|^2}]$.

This claim follows from the following three assertions.

♠ $\max\{\gamma_{\min}, [-\beta'_\mathrm{P}]_+\} = \gamma_{\min}$: If $\beta'_\mathrm{P} \geq 0$, then this assertion vacuously holds since $\gamma_{\min} \geq 0$. Let $\operatorname{rank} L < n$ and $\beta'_\mathrm{P} = \rho_A < 0$, so that by assumption $\rho_B > 0$ and $\beta_\mathrm{P} < 0$. Consider the following two cases.

◇ $\beta_\mathrm{D} \leq 0$: It holds that $\delta = 1$ and thus

$$\gamma_{\min} = \frac{-2\beta_\mathrm{P}}{\delta + \sqrt{\delta^2 - 4\beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2}} = \frac{-2\beta_\mathrm{P}}{1 + \sqrt{1 - 4\beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2}} \geq -\beta_\mathrm{P} \overset{\text{(B.9)}}{\geq} -\rho_A = [-\beta'_\mathrm{P}]_+. \tag{B.10}$$

◇ $\beta_\mathrm{D} > 0$: By definition $\mu_A > 0$ and $\mu_B > 0$. Furthermore,

$$\gamma_{\min} > [-\beta'_\mathrm{P}]_+ \iff \frac{1}{\gamma_{\min}} < \frac{1}{[-\beta'_\mathrm{P}]_+} \iff \frac{\delta + \sqrt{\delta^2 - 4\beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2}}{-2\beta_\mathrm{P}} < -\frac{1}{\rho_A} \iff \frac{\beta_\mathrm{P}}{\rho_A} - \frac{1}{2}\delta > \frac{1}{2}\sqrt{\delta - 4\beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2} \tag{B.11}$$

$$\overset{\text{(a)}}{\iff} \frac{\beta_\mathrm{P}^2}{\rho_A^2} - \frac{\beta_\mathrm{P}}{\rho_A}\delta + \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2 > 0 \overset{\text{(b)}}{\iff} \mu_B\underbrace{(1 - \mu_A\rho_A\|L\|^2)}_{=:\xi_A} + \mu_A\underbrace{(1 - \mu_B\rho_B\sigma_d^2)}_{=:\xi_B} > 0, \tag{B.12}$$

where (a) holds since $\frac{\beta_\mathrm{P}}{\rho_A} - \frac{1}{2}\delta > 1 - \frac{1}{2}(1 + \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2) \geq \frac{1}{2}$, and it is obtained by squaring both sides of the inequality, while in (b) both sides of the inequality were multiplied by $-\frac{1}{\beta_\mathrm{P}}(\mu_A + \mu_B)(\rho_A + \rho_B) > 0$, and $\delta = 1 + \beta_\mathrm{D}\beta_\mathrm{P}(\|L\|^2 - \sigma_d^2)$, $\beta_\mathrm{D} = \mu_A\mu_B/(\mu_A+\mu_B)$ and $\beta_\mathrm{P} = \rho_A\rho_B/(\rho_A+\rho_B)$ were used. Since $\mu_A\rho_A < 0$, it holds that $\xi_A > 0$, while Lemma A.4 guarantees that $\xi_B > 0$, completing the proof.

♠ $\min\{\gamma_{\max}, \frac{1}{[-\beta'_\mathrm{D}]_+\|L\|^2}\} = \gamma_{\max}$: Since

$$\gamma_{\max} < \frac{1}{[-\beta'_\mathrm{D}]_+\|L\|^2} \iff \frac{1}{\gamma_{\max}\|L\|^2} > [-\beta'_\mathrm{D}]_+ \iff \frac{2[-\beta_\mathrm{D}]_+}{\delta + \sqrt{\delta^2 - 4\beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2}} > [-\beta'_\mathrm{D}]_+$$

it follows that verifying this assertion is analogous to verifying the inequality $\gamma_{\min} > [-\beta'_\mathrm{P}]_+$ from (B.12), swapping the roles of $\beta_\mathrm{P}, \beta'_\mathrm{P}$ with $\beta_\mathrm{D}, \beta'_\mathrm{D}$.

♠ $\max\{\tau_{\min}(\gamma), [-\beta'_\mathrm{D}]_+\} = \tau_{\min}(\gamma)$ for all $\gamma \in (\gamma_{\min}, \gamma_{\max})$: If $\beta'_\mathrm{D} \geq 0$, then this assertion vacuously holds since $\tau_{\min}(\gamma) \geq 0$ for all $\gamma > 0$. Let $\operatorname{rank} L < m$ and $\beta'_\mathrm{D} = \mu_B < 0$. Then, by assumption $\mu_A > 0$ and $\beta_\mathrm{D} < 0$. We consider the following two cases.

◇ $\beta_\mathrm{P} \leq 0$: Owing to (B.10) it holds that $\gamma + \beta_\mathrm{D} > 0$. Consequently, using that $\delta = 1$ follows that

$$\tau_{\min}(\gamma) = \frac{-\beta_\mathrm{D}(\gamma + \beta_\mathrm{P})}{\gamma(\delta - \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2) + \beta_\mathrm{P}} = \frac{-\beta_\mathrm{D}(\gamma + \beta_\mathrm{P})}{\gamma(1 - \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2) + \beta_\mathrm{P}} \geq -\beta_\mathrm{D} \overset{\text{(B.9)}}{\geq} -\mu_B = [-\beta'_\mathrm{D}]_+,$$

where we used that $1 - \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2 \in (3/4, 1]$ owing to Assumption III*(iv)*.

◇ $\beta_\mathrm{P} > 0$: By definition $\gamma(\delta - \beta_\mathrm{P}\beta_\mathrm{D}\|L\|^2) + \beta_\mathrm{P} = \gamma(1 - \beta_\mathrm{P}\beta_\mathrm{D}\sigma_d^2) + \beta_\mathrm{P} > 0$ and thus it holds that

$$\begin{aligned}\tau_{\min}(\gamma) > [-\beta'_\mathrm{D}]_+ &\iff \left(\gamma(1 - \beta_\mathrm{P}\beta_\mathrm{D}\sigma_d^2) + \beta_\mathrm{P}\right)\mu_B > \beta_\mathrm{D}(\gamma + \beta_\mathrm{P}) \\ &\iff 0 < \gamma\left(\rho_A + \rho_B(1 - \mu_A\rho_A\sigma_d^2)\right) + \rho_A\rho_B\end{aligned} \tag{B.13}$$

where we multiplied by $\frac{1}{\mu_B^2}(\mu_A + \mu_B)(\rho_A + \rho_B) > 0$ to obtain the final equivalence. Since $\beta_\mathrm{P} > 0$, it follows that $\rho_A > 0$ and $\rho_B > 0$ and thus satisfaction of (B.13) is guaranteed through Lemma A.4.

Having established that the conditions from Stepsize rule I reduce to $\gamma \in (\gamma_{\min}, \gamma_{\max})$ and $\tau \in \left(\tau_{\min}(\gamma), \frac{1}{\gamma\|L\|^2}\right]$, it is straightforward to verify that these intervals match the ones provided in Table 2. 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 Table 3 is obtained by plugging in β_{P}' and β_{D}' from (5.3) into Table 1. Moreover, as a consequence of (B.9), it holds that $\min\left\{1 + \frac{1}{\gamma}\beta_{\mathrm{P}}, 1 + \frac{1}{\tau}\beta_{\mathrm{D}}\right\} \leq \min\left\{1 + \frac{1}{\gamma}\rho_A, 1 + \frac{1}{\tau}\mu_B\right\} \leq \eta'$ and the following assertions hold.

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

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

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

The claimed equivalence follows immediately from these two assertions, completing the proof. $\square$

References

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

[2] A. Auslender and M. Teboulle, *Asymptotic cones and functions in optimization and variational inequalities*, Springer Science & Business Media, Berlin, 2006.

[3] J. K. Baksalary, F. Pukelsheim, and G. P. Styan, *Some properties of matrix partial orderings*, Linear Algebra and its Applications, 119 (1989), pp. 57–85.

[4] V. Balakrishnan and L. Vandenberghe, *Semidefinite programming duality and linear time-invariant systems*, IEEE Transactions on Automatic Control, 48 (2003), pp. 30–41.

[5] S. Banert, M. Upadhyaya, and P. Giselsson, *The Chambolle–Pock method converges weakly with $\theta > 1/2$ and $\tau\sigma\|l\|^2 < 4/(1 + 2\theta)$*, 2023, arXiv preprint arXiv:2309.03998.

[6] H. H. Bauschke, W. M. Moursi, and X. Wang, *Generalized monotone operators and their averaged resolvents*, Mathematical Programming, 189 (2021), pp. 55–74.

[7] D. S. Bernstein, *Matrix Mathematics*, Princeton University Press, Princeton, 2009.

[8] R. I. Boţ and D.-K. Nguyen, *The proximal alternating direction method of multipliers in the nonconvex setting: Convergence analysis and rates*, Mathematics of Operations Research, 45 (2020), pp. 682–712.

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

[10] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, SIAM, 1994.

[11] K. Bredies, E. Chenchene, D. A. Lorenz, and E. Naldi, *Degenerate preconditioned proximal point algorithms*, SIAM Journal on Optimization, 32 (2022), pp. 2376–2401.

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

[13] P. L. Combettes and T. Pennanen, *Proximal methods for cohypomonotone operators*, SIAM Journal on Control and Optimization, 43 (2004), pp. 731–742.

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

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

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

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

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

[19] B. Evens, P. Pas, P. Latafat, and P. Patrinos, *Convergence of the preconditioned proximal point method and Douglas-Rachford splitting in the absence of monotonicity*, 2023, arXiv preprint arXiv:2305.03605.

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

[21] G. H. Golub and C. F. Van Loan, *Matrix Computations*, JHU press, 2013.

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

[23] J. Grosz, *A note on the general hermitian solution to AXA*= B*, Bulletin of the Malaysian Mathematical Sciences Society, 21 (1998).

[24] N. Hallak and M. Teboulle, *An adaptive Lagrangian-based scheme for nonconvex composite optimization*, Mathematics of Operations Research, (2023).

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

[26] A. Helmersson, *Methods for robust gain scheduling*, PhD thesis, Linköping University Electronic Press, 1995.

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

[28] C. G. Khatri and S. K. Mitra, *Hermitian and nonnegative definite solutions of linear matrix equations*, SIAM Journal on Applied Mathematics, 31 (1976), pp. 579–585.

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

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

[31] G. Li and T. Pong, *Global convergence of splitting methods for nonconvex composite optimization*, SIAM Journal on Optimization, 25 (2015), pp. 2434–2460.

[32] Y. Li and M. Yan, *On the improved conditions for some primal-dual algorithms*, 2022, arXiv preprint arXiv:2201.00139.

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

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

[35] T. J. Meijer, T. Holicki, S. J. A. M. van den Eijnden, C. W. Scherer, and W. P. M. H. Heemels, *The non-strict projection lemma*, 2023, arXiv preprint arXiv:2305.08735.

[36] G. J. Minty, *Monotone (nonlinear) operators in Hilbert space*, Duke Mathematical Journal, 29 (1962), pp. 341 – 346.

[37] S. K. Mitra, P. Bhimasankaram, and S. B. Malik, *Matrix partial orders, shorted operators and applications*, vol. 10, World Scientific, 2010.

[38] T. Möllenhoff, E. Strekalovskiy, M. Moeller, and D. Cremers, *The primal-dual hybrid gradient method for semiconvex splittings*, SIAM Journal on Imaging Sciences, 8 (2015), pp. 827–857.

[39] D. O'Connor and L. Vandenberghe, *On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting*, Mathematical Programming, 179 (2020), pp. 85–108.

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

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

[42] R. Penrose, *A generalized inverse for matrices*, in Mathematical proceedings of the Cambridge philosophical society, vol. 51, Cambridge University Press, 1955, pp. 406–413.

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

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

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

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

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

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

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

[50] R. Shefi and M. Teboulle, *Rate of convergence analysis of decomposition methods based on the proximal method of multipliers for convex minimization*, SIAM Journal on Optimization, 24 (2014), pp. 269–297.

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

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

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

[54] Y. Tian, *Equalities and inequalities for inertias of Hermitian matrices with applications*, Linear algebra and its applications, 433 (2010), pp. 263–296.

[55] Y. Tian, *Analytical solutions to some optimization problems on ranks and inertias of matrix-valued functions subject to linear matrix inequalities*, 2013, arXiv preprint arXiv:1301.0986.

[56] C. Yonglin and Z. Bingjun, *On g-inverses and the nonsingularity of a bordered matrix*, Linear Algebra and its Applications, 133 (1990), pp. 133–151.

[57] M. Zhu and T. Chan, *An efficient primal-dual hybrid gradient algorithm for total variation image restoration*, Ucla Cam Report, 34 (2008), pp. 8–34.