

Local linear convergence of the primal–dual hybrid gradient algorithm for semidefinite programming

Xin Jiang*

July 8, 2026

Abstract

Primal–dual first-order methods are widely used for large-scale semidefinite programming (SDP), but their ability to compute highly accurate solutions is not well explained by global convergence theory alone. We study the local convergence of the primal–dual hybrid gradient (PDHG) method applied to a standard primal–dual SDP pair. We show that PDHG converges eventually (R -)linearly whenever the limiting KKT point satisfies either strict complementarity or primal–dual nondegeneracy. The proof views PDHG as a preconditioned proximal point method for the KKT inclusion and combines its descent inequality with a local error bound. Under strict complementarity, the error bound follows from the local spectral geometry of the positive semidefinite cone; under primal–dual nondegeneracy, it follows from strong regularity of the KKT mapping. We also give a simple SDP instance where both regularity conditions fail and PDHG can converge only sublinearly. This contrasts with linear programming, where PDHG admits a local linear convergence regime even for degenerate instances. Numerical experiments support the theory and identify difficult SDP instances where PDHG struggles to reach high accuracy.

1 Introduction

Semidefinite programming is a central model in convex optimization. It combines affine constraints with the geometry of the positive semidefinite (PSD) cone, and arises in control, signal processing, statistics, convex relaxations of polynomial optimization and combinatorial optimization, and many other fields [2, 9, 11, 38, 46]. The growing demand for solving large-scale instances has motivated the use of primal–dual first-order methods, whose main operations typically are applications of linear maps and projections onto the PSD cone. This paper develops a local convergence theory for one such method, the primal–dual hybrid gradient (PDHG) algorithm, applied to semidefinite programming.

We consider the primal–dual pair of semidefinite programs (SDPs)

$$\begin{array}{ll} \text{Primal:} & \begin{array}{l} \text{minimize} \quad \langle C, X \rangle \\ \text{subject to} \quad \mathcal{A}X = b \\ X \in \mathbb{S}_+^n \end{array} & \text{Dual:} & \begin{array}{l} \text{maximize} \quad b^\top y \\ \text{subject to} \quad \mathcal{A}^*y + S = C \\ S \in \mathbb{S}_+^n \end{array} \end{array} \quad (1)$$

with primal variable $X \in \mathbb{S}^n$ and dual variables $S \in \mathbb{S}^n$, $y \in \mathbb{R}^m$. Here $\mathbb{S}^n$ is the set of real symmetric $n \times n$ matrices, and $\mathbb{S}_+^n$ is the set of positive semidefinite matrices in $\mathbb{S}^n$. The linear operator $\mathcal{A}: \mathbb{S}^n \rightarrow \mathbb{R}^m$ is defined by $\mathcal{A}X = (\langle A_1, X \rangle, \dots, \langle A_m, X \rangle)$, and $\mathcal{A}^*y = \sum_{i=1}^m y_i A_i$ is its adjoint operator. The coefficients $C, A_1, \dots, A_m$ are symmetric $n \times n$ matrices.

When applied to (1), PDHG [14, 24] with stepsizes $\tau, \sigma > 0$ takes the form

$$X_{k+1} = \Pi_{\mathbb{S}_+^n}(X_k - \tau(C - \mathcal{A}^*y_k)) \quad (2a)$$

$$y_{k+1} = y_k + \sigma(b - \mathcal{A}(2X_{k+1} - X_k)). \quad (2b)$$

*Department of Industrial and Systems Engineering, University of Houston. Email: xinjiang@uh.edu

The stepsizes are fixed throughout the paper and satisfy

$$\tau > 0, \quad \sigma > 0, \quad \tau\sigma\|\mathcal{A}\|^2 < 1, \quad (3)$$

where $\|\mathcal{A}\|$ is the operator norm induced by the Frobenius norm on $\mathbb{S}^n$ and the Euclidean norm on $\mathbb{R}^m$. Each iteration consists of one projection onto the PSD cone and two applications of the linear maps $\mathcal{A}$ and $\mathcal{A}^*$. This simple structure makes PDHG attractive for large-scale SDPs [42, 60]. At the same time, practitioners often view primal–dual first-order methods as effective for finding moderate-accuracy SDP solutions, but less reliable when high accuracy is required. This view is consistent with the classic global convergence theory: sublinear complexity bounds suggest that progress should slow as the iterates approach the solution [14, 15].

Recent results have begun to refine this picture. For linear programming, a special case of SDP, [44] identifies a two-phase phenomenon for PDHG: after an initial global phase, the iterates enter a local phase governed by the active geometry of the optimal face and then converge at a (R -)linear rate. For semidefinite programming, [28, 35] provide related local (R -)linear convergence results for the alternating direction method of multipliers (ADMM) under different regularity assumptions at the limit point. These developments suggest that the eventual behavior of a first-order method is determined not only by the algorithm, but also by the variational geometry of the limiting KKT point.

This leaves open the corresponding local theory for PDHG on SDP. The LP results rely on polyhedral geometry, while the ADMM results use the structure of the ADMM operator. Thus, for the nonpolyhedral PSD cone and the PDHG iteration, local R -linear convergence remains unresolved in general. We develop this local theory from the monotone inclusion formulation of the KKT system for (1). Specifically, we view PDHG as a preconditioned proximal point method applied to the KKT inclusion [32]. This viewpoint yields a descent inequality in a problem-dependent metric, and the remaining task is to obtain a compatible local error bound.

We obtain such error bounds under either of two classical regularity conditions: *strict complementarity* and *primal–dual nondegeneracy* [3]. The former is spectral: the positive eigenspaces of the primal matrix and the dual slack complement each other, leaving no common degenerate zero eigenspace. The latter is variational: the affine constraints meet the tangent geometry of the PSD cone without hidden flat directions, on both the primal and dual sides. These two regularity conditions also play rather different roles in practice. Strict complementarity can often be diagnosed from a computed primal–dual solution by inspecting the spectra of the primal matrix and the dual slack. Primal–dual nondegeneracy, in contrast, requires checking a larger and more delicate linear-algebraic condition and can be expensive for large-scale SDPs.

In view of their different geometric meanings and practical verifiability, we treat strict complementarity and primal–dual nondegeneracy as separate SDP regularity mechanisms for local convergence. We show that either mechanism is sufficient for local R -linear convergence of PDHG. We also present an SDP instance suggesting that, without either condition, PDHG may fail to enter a local linear regime.

Contributions. The contributions of this paper are twofold.

- *Theoretical results.* We prove that PDHG is eventually (R -)linearly convergent for SDP if the limiting KKT point satisfies either strict complementarity or primal–dual nondegeneracy. The proof uses the preconditioned proximal-point interpretation of PDHG and shows that each regularity condition provides the local error bound needed for the descent argument. The two cases are different in nature: strict complementarity gives an error bound through the local spectral geometry of the PSD cone, whereas primal–dual nondegeneracy gives an error bound through variational regularity of the KKT mapping. Under strict complementarity, the same spectral analysis also yields finite-time rank identification of the primal iterates; this conclusion is consistent with the more general partial smoothness theory [22, 39, 63]. We also present a simple SDP instance where both strict complementarity and primal–dual nondegeneracy fail. A formal local reduction of the PDHG map, supported by numerical evidence, predicts sublinear decay. This provides evidence that, unlike in LP, local linear convergence of PDHG on SDP cannot be expected without additional regularity.

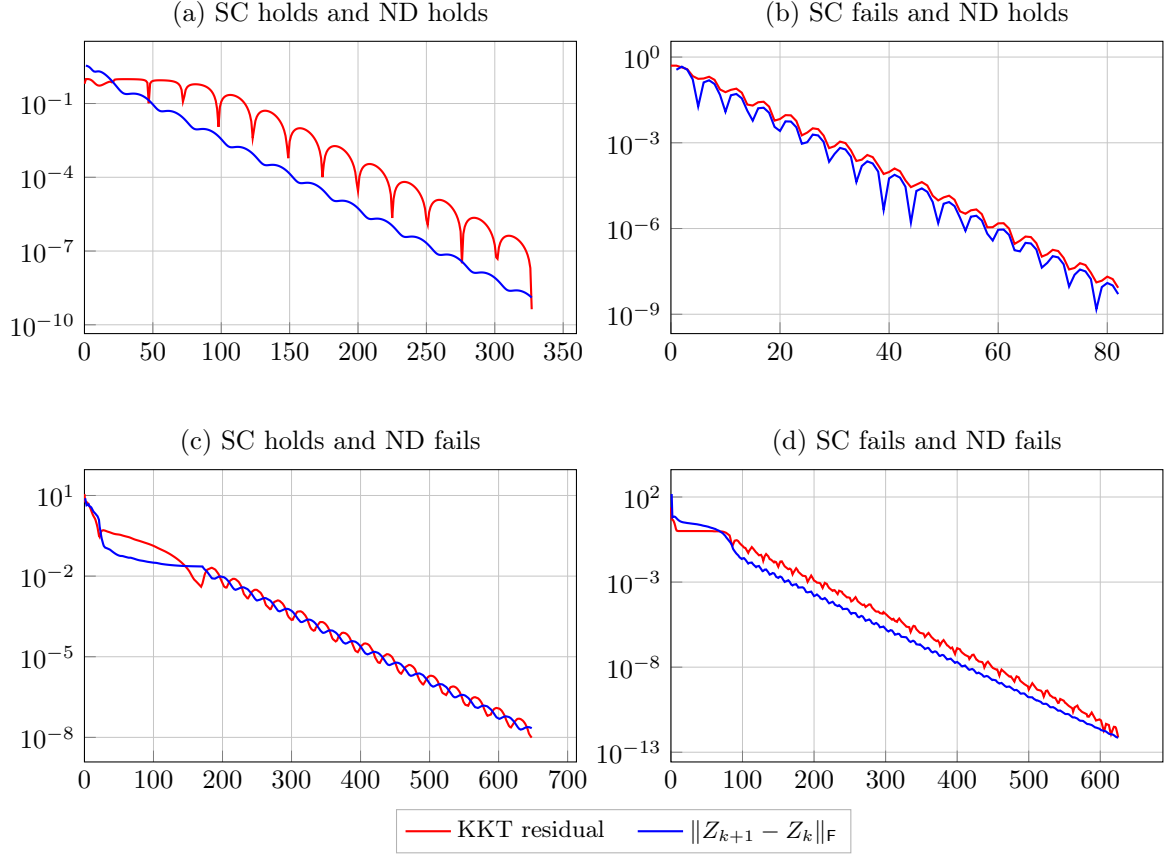


Figure 1: Convergence behavior on four simple SDP instances with different combinations of strict complementarity (SC) and nondegeneracy (ND) conditions. (a) A toy structure-from-motion problem from [29]. (b) A toy example from [62, p. 44]. (c) The second-order relaxation of a polynomial optimization problem [59, 64]. (d) A constructed toy example with a unique KKT point that violates dual nondegeneracy and strict complementarity. The red curve is the KKT residual, and the blue curve is the fixed-point difference $\|Z_{k+1} - Z_k\|_F$, where $Z_k = X_k - \tau(C - \mathcal{A}^*y_k)$. In all four instances, PDHG with fixed stepsizes eventually exhibits local linear convergence. Additional numerical results using real-world data are presented in Section 7.

- *Empirical results.* We complement the theory with extensive numerical experiments. The experiments illustrate the predicted local linear convergence on SDP instances with different combinations of strict complementarity and primal–dual nondegeneracy. Figure 1 gives a representative summary: in these examples, PDHG enters a clear linear convergence regime across all four regularity profiles. The full set of experiments, reported in Section 7, covers a broad range of SDP instances.

We also document instances for which PDHG does not reach high accuracy within a prescribed computational budget of 10^5 iterations; in these cases, the overall KKT residual remains larger than 10^{-8} . A common feature of these difficult instances is that the minimum positive eigenvalues of the final primal and dual slack iterates are close to zero but not numerically zero. This behavior closely resembles difficult cases observed for LP [44].

Outline. The rest of the paper is organized as follows. After reviewing related work in Section 2, we introduce the problem setting and the two regularity conditions used in the analysis: strict complementarity

and primal–dual nondegeneracy. [Sections 4](#) and [5](#) establish local linear convergence of PDHG for SDP under these two regularity conditions, respectively. [Section 6](#) discusses the scope and limitations of the results. It gives a simple example in which a formal local reduction of the PDHG map predicts sublinear convergence, compares the proof mechanisms with existing analyses of PDHG for LP and ADMM for SDP, and gives a direct proof of finite-time rank identification for PDHG on SDP. [Section 7](#) reports numerical evidence for local linear convergence and [Section 8](#) concludes the paper.

2 Related work

Algorithms for semidefinite programming. Semidefinite programming has traditionally been solved to high accuracy by primal–dual interior-point methods; see, *e.g.*, [\[54, 58, 62\]](#). These methods exploit the self-dual structure of the positive semidefinite cone and have a mature theory and reliable implementations, including `SeDuMi` and `SDPT3` [\[51, 55\]](#). Their per-iteration cost, however, limits their use on very large instances. This has motivated several complementary directions, including chordal and sparse matrix cone techniques [\[57\]](#), augmented-Lagrangian, ADMM, and semismooth Newton methods [\[27, 61, 65, 66\]](#), low-rank factorization methods [\[12, 13\]](#), and more recent low-rank splitting schemes [\[1, 20, 30, 47, 53\]](#). The present paper is not concerned with designing a new SDP solver. Its goal is instead to understand the local behavior of a basic primal–dual first-order method when it is applied directly to the standard primal–dual SDP pair.

PDHG. The primal–dual hybrid gradient (PDHG) algorithm, also known as the Chambolle–Pock method, is a primal–dual splitting method for saddle-point problems and monotone inclusions [\[14, 15, 24, 49\]](#). Its ergodic convergence is well understood, and related primal–dual first-order methods have also been developed for sparse SDP [\[33\]](#). When [\(3\)](#) holds, PDHG can be interpreted as a preconditioned proximal point method for the KKT inclusion [\[32, 34, 48\]](#). This interpretation is central to our analysis.

PDHG for LP. In linear programming, PDHG has recently become a practical large-scale method through the PDLP line of work, where diagonal scaling, restarts, presolve, and other implementation techniques make it competitive on very large instances [\[4–6, 18, 40, 42, 43\]](#). These developments have also stimulated refined theoretical analyses of PDHG for LP, including two-stage behavior based on finite basis identification followed by local linear convergence [\[44\]](#). SDP differs from LP in an essential way: the PSD cone is nonpolyhedral, and the projection is not piecewise affine near matrices with zero eigenvalues. Recent work has also considered adaptive primal–dual splitting methods designed specifically for large-scale SDP [\[60\]](#).

Local linear convergence and variational regularity. Local linear convergence of proximal splitting methods is often proved through error bounds, metric subregularity, partial smoothness, or local linearization after active-manifold identification. For the proximal point method, this viewpoint goes back to Rockafellar [\[50\]](#); for Douglas–Rachford splitting and ADMM, related mechanisms appear in [\[10, 19, 23, 25, 26, 41\]](#). In SDP, strict complementarity and primal–dual nondegeneracy play analogous (yet different) regularizing roles. The nondegeneracy conditions characterize how the affine constraints meet the local geometry of the semidefinite cone [\[3\]](#) and are later related to strong regularity of the KKT system [\[16\]](#). These variational properties are exploited in [Section 5](#) to establish the local linear convergence of PDHG when applied to SDP.

Relation to ADMM analyses for SDP. The closest work to the present paper is the local convergence theory of ADMM for SDP under nondegeneracy or strict complementarity [\[28, 35\]](#), together with related second-order studies of ADMM dynamics [\[36\]](#). Our analysis is different in both the algorithmic map and the metric. ADMM is naturally studied through the Douglas–Rachford operator in a signed matrix variable, whereas PDHG is the P -resolvent of the reduced KKT mapping in the primal–dual variable (X, y) . Under strict complementarity, we exploit the smoothness of the PSD-cone projection near the limiting signed matrix and prove contraction normal to the local KKT manifold. Under primal–dual nondegeneracy, we use strong

metric subregularity of the KKT mapping to obtain local Q -linear convergence directly from the descent inequality. More detailed discussion can be found in [Section 6.2](#).

3 Problem setting and regularity

We now fix the notation and regularity assumptions used throughout the analysis. The two conditions of interest, strict complementarity and primal–dual nondegeneracy, capture different ways in which the affine constraints interact with the geometry of the positive semidefinite cone. The first condition gives a smooth local model of the cone projection, whereas the second gives an error bound for the KKT mapping. These two mechanisms lead to the two convergence arguments developed in the next sections.

3.1 Notation

Let $\mathbb{S}^n$ denote the set of real symmetric $n \times n$ matrices, equipped with the inner product $\langle X, Y \rangle = \text{tr}(XY)$ and the Frobenius norm $\|X\|_F$. The cone of positive semidefinite matrices in $\mathbb{S}^n$ is denoted by $\mathbb{S}_+^n$. We write $X \succeq 0$ and $X \succ 0$ to mean that X is positive semidefinite and positive definite, respectively. For a linear operator A , its range and nullspace are denoted by $\text{Range}(A)$ and $\text{Null}(A)$, respectively. The rank and inertia of $X \in \mathbb{S}^n$ are denoted by $\text{rank } X$ and $\text{inertia}(X)$, respectively. For a set-valued mapping F , its fixed-point set is denoted by $\text{Fix}(F) = \{u \mid F(u) = u\}$. Let e_i denote the i th unit vector. Define $E_{ii} = e_i e_i^\top$ and $E_{ij} = (e_i e_j^\top + e_j e_i^\top)/2$, where the ambient dimension is understood from context. We denote $[n] = \{1, \dots, n\}$, $\mathbf{diag}(x)$ for the diagonal matrix with diagonal x , and $\circ$ for the Hadamard product; *i.e.*, $(A \circ B)_{ij} = A_{ij} B_{ij}$. For a linear subspace $\mathcal{L}$, denote by $\mathcal{L}^\perp$ its orthogonal complement.

For a set $\mathcal{C}$, we denote by $\mathbf{aff } \mathcal{C}$ its affine hull. If $\mathcal{C}$ is nonempty and closed, then $\Pi_{\mathcal{C}}$ denotes the Euclidean projection onto $\mathcal{C}$ and $\mathbf{dist}(x, \mathcal{C})$ denotes the Euclidean distance from x to $\mathcal{C}$:

$$\Pi_{\mathcal{C}}(x) = \underset{y \in \mathcal{C}}{\operatorname{argmin}} \|x - y\|, \quad \mathbf{dist}(x, \mathcal{C}) = \inf_{y \in \mathcal{C}} \|x - y\|.$$

If moreover $\mathcal{C}$ is convex, then the normal cone to $\mathcal{C}$ at x is $N_{\mathcal{C}}(x) = \{v \mid \langle v, y - x \rangle \leq 0 \text{ for all } y \in \mathcal{C}\}$. The closed Euclidean ball of radius ϵ centered at x is denoted by $\mathcal{B}_\epsilon(x)$.

For a positive definite self-adjoint operator P , define $\langle u, v \rangle_P = \langle u, Pv \rangle$ and $\|u\|_P = \sqrt{\langle u, u \rangle_P}$. We write $u \perp_P v$ if $\langle u, v \rangle_P = 0$. For a nonempty closed convex set $\mathcal{C}$, $\Pi_{\mathcal{C}}^P$ denotes the projection onto $\mathcal{C}$ in the P -metric and $\mathbf{dist}_P(u, \mathcal{C})$ denotes the corresponding distance:

$$\Pi_{\mathcal{C}}^P(u) = \underset{v \in \mathcal{C}}{\operatorname{argmin}} \|u - v\|_P, \quad \mathbf{dist}_P(u, \mathcal{C}) = \inf_{v \in \mathcal{C}} \|u - v\|_P.$$

We denote by $\mathcal{B}_\epsilon^P(u)$ the closed P -metric ball of radius ϵ centered at u .

3.2 Standing assumptions

We make the following assumptions on [\(1\)](#).

Assumption 1. *The linear operator $\mathcal{A}: \mathbb{S}^n \rightarrow \mathbb{R}^m$ is surjective, and the KKT set*

$$\Omega_\star = \{(X, y, S) \in \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n \mid \mathcal{A}X = b, X \in \mathbb{S}_+^n, \mathcal{A}^*y + S = C, S \in \mathbb{S}_+^n, \langle X, S \rangle = 0\}$$

is nonempty.

Under [Assumption 1](#), any KKT point $(X_\star, y_\star, S_\star)$ satisfies complementary slackness. Hence $X_\star$ and $S_\star$ admit a simultaneous orthogonal decomposition

$$\begin{aligned} X_\star &= Q_\star \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix} Q_\star^\top, & \Lambda_X &= \mathbf{diag}(\lambda_1, \dots, \lambda_r), \\ S_\star &= Q_\star \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_S \end{bmatrix} Q_\star^\top, & \Lambda_S &= -\mathbf{diag}(\lambda_{n-s+1}, \dots, \lambda_n), \end{aligned} \tag{4}$$

where $Q_\star \in \mathbb{R}^{n \times n}$ is orthogonal and $\lambda_1 \geq \dots \geq \lambda_r > 0 > \lambda_{n-s+1} \geq \dots \geq \lambda_n$, with $r + s \leq n$. Consequently, $\text{rank } X_\star + \text{rank } S_\star \leq n$.

3.3 Strict complementarity

A KKT point $(X_\star, y_\star, S_\star)$ is *strictly complementary* if $X_\star + S_\star \succ 0$, or equivalently in (4),

$$\text{rank } X_\star + \text{rank } S_\star = n. \quad (5)$$

Strict complementarity separates the positive eigenspaces of $X_\star$ and $S_\star$. It rules out directions in which both the primal and dual slack matrices are degenerate, and hence makes the active rank pattern stable under small perturbations. This is the mechanism that turns the local geometry of the positive semidefinite cone into a smooth one. In Section 4, this separation gives a smooth local model of $\Pi_{\mathbb{S}_+^n}$ near $X_\star - \tau S_\star$ and fixes the local face structure of the KKT set. The assumption needed in Section 4 is formalized below.

Assumption 2. *The PDHG iteration (2) converges to a KKT point $(X_\star, y_\star, S_\star)$ satisfying strict complementarity (5).*

Assumption 2 is a mild assumption in the sense that the existence of a strictly complementary solution is a generic property of SDPs [3, Theorem 15].

3.4 Primal–dual nondegeneracy

Another useful regularity condition is primal and dual nondegeneracy, which is related to the four important subspaces [3]:

$$\begin{aligned} \mathcal{T}_{X_\star} &= \left\{ Q_\star \begin{bmatrix} H_{11} & H_{21}^\top \\ H_{21} & 0 \end{bmatrix} Q_\star^\top \mid H_{11} \in \mathbb{S}^r, H_{21} \in \mathbb{R}^{(n-r) \times r} \right\} \\ \mathcal{N}_{X_\star} = \mathcal{T}_{X_\star}^\perp &= \left\{ Q_\star \begin{bmatrix} 0 & 0 \\ 0 & H_{22} \end{bmatrix} Q_\star^\top \mid H_{22} \in \mathbb{S}^{n-r} \right\} \\ \mathcal{T}_{S_\star} &= \left\{ Q_\star \begin{bmatrix} 0 & H_{21}^\top \\ H_{21} & H_{22} \end{bmatrix} Q_\star^\top \mid H_{22} \in \mathbb{S}^s, H_{21} \in \mathbb{R}^{s \times (n-s)} \right\} \\ \mathcal{N}_{S_\star} = \mathcal{T}_{S_\star}^\perp &= \left\{ Q_\star \begin{bmatrix} H_{11} & 0 \\ 0 & 0 \end{bmatrix} Q_\star^\top \mid H_{11} \in \mathbb{S}^{n-s} \right\}. \end{aligned}$$

Here $\mathcal{T}_{X_\star}$ and $\mathcal{T}_{S_\star}$ are the lineality spaces of the tangent cones to $\mathbb{S}_+^n$ at $X_\star$ and $S_\star$, respectively, and $\mathcal{N}_{X_\star}$ and $\mathcal{N}_{S_\star}$ are their orthogonal complements.

A KKT point $(X_\star, y_\star, S_\star)$ is said to satisfy *primal–dual nondegeneracy* if

$$\begin{aligned} \text{Primal nondegeneracy: } \mathcal{T}_{X_\star} + \text{Null}(\mathcal{A}) &= \mathbb{S}^n & \iff & \mathcal{N}_{X_\star} \cap \text{Range}(\mathcal{A}^*) = \{0\} \\ \text{Dual nondegeneracy: } \mathcal{T}_{S_\star} + \text{Range}(\mathcal{A}^*) &= \mathbb{S}^n & \iff & \mathcal{N}_{S_\star} \cap \text{Null}(\mathcal{A}) = \{0\}. \end{aligned} \quad (6)$$

These conditions are nondegeneracy conditions on the way the affine constraints meet the local geometry of the cone. Primal nondegeneracy says that the feasible affine directions, together with the tangent directions of the cone at $X_\star$, span the whole space. Dual nondegeneracy gives the analogous condition for the dual slack matrix $S_\star$.

In general, primal–dual nondegeneracy is independent of strict complementarity. When strict complementarity holds, then the two conditions have a simple interpretation: primal (resp., dual) nondegeneracy is equivalent to uniqueness of the dual (resp., primal) optimal solution [3]. Moreover, primal–dual nondegeneracy provides an error bound for the KKT mapping and thus gives a direct route to local linear convergence of PDHG in Section 5, without constructing a smooth local model of the PSD-cone projection.

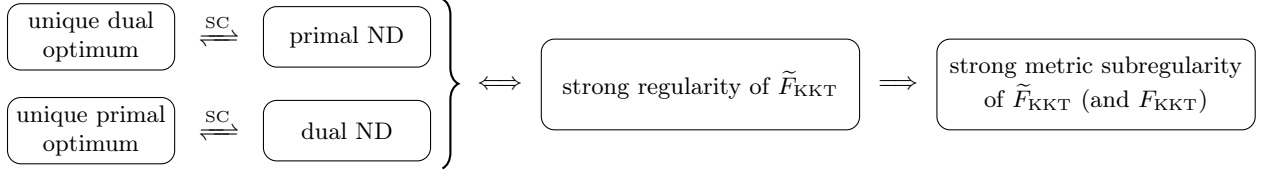


Figure 2: Relationships among strict complementarity (SC), primal/dual nondegeneracy (ND), primal/dual optima, and regularity properties of the two KKT mappings $\tilde{F}_{\text{KKT}}$ and F_{KKT} .

To make this error-bound statement precise, we introduce two KKT mappings associated with (1). The extended KKT mapping is

$$\tilde{F}_{\text{KKT}} = \begin{bmatrix} \mathcal{A}X - b \\ \mathcal{A}^*y + S - C \\ S + N_{\mathbb{S}_+^n}(X) \end{bmatrix}.$$

Thus $0 \in \tilde{F}_{\text{KKT}}(X, y, S)$ is equivalent to primal feasibility, dual feasibility, and complementarity. We also define the reduced KKT mapping, obtained by eliminating $S = C - \mathcal{A}^*y$,

$$F_{\text{KKT}} = \begin{bmatrix} C - \mathcal{A}^*y + N_{\mathbb{S}_+^n}(X) \\ \mathcal{A}X - b \end{bmatrix}.$$

By [16, Theorem 18], primal–dual nondegeneracy at a KKT point (X_*, y_*, S_*) is equivalent to *strong regularity* of the monotone inclusion $0 \in \tilde{F}_{\text{KKT}}(X, y, S)$ at (X_*, y_*, S_*) . We use this equivalence only through one of its standard consequences: strong regularity implies *strong metric subregularity* [21, Theorems 4.2 and 5.2]. That is, there exist constants $\kappa > 0$ and $\epsilon > 0$ such that

$$\|(X, y, S) - (X_*, y_*, S_*)\| \leq \kappa \mathbf{dist}(0, \tilde{F}_{\text{KKT}}(X, y, S)) \quad \text{whenever } \|(X, y, S) - (X_*, y_*, S_*)\| \leq \epsilon.$$

This estimate differs from ordinary metric subregularity in that the left-hand side is the distance to the particular KKT point, rather than the distance to the solution set Ω_* .

The same property holds for the reduced KKT mapping. Indeed, set $S = C - \mathcal{A}^*y$. Then the second component of $\tilde{F}_{\text{KKT}}(X, y, S)$ vanishes and

$$\mathbf{dist}(0, \tilde{F}_{\text{KKT}}(X, y, C - \mathcal{A}^*y)) \leq \kappa_1 \mathbf{dist}(0, F_{\text{KKT}}(X, y)).$$

Moreover, since $S - S_* = -\mathcal{A}^*(y - y_*)$,

$$\|(X, y) - (X_*, y_*)\| \leq \kappa_2 \|(X, y, S) - (X_*, y_*, S_*)\|.$$

Combining these inequalities yields strong metric subregularity of F_{KKT} at $(z_*, 0)$, where $z_* = (X_*, y_*)$. Equivalently, for any positive definite metric P , there exist constants $\mu > 0$ and $\delta > 0$ such that

$$\|z - z_*\|_P \leq \mu \mathbf{dist}_{P^{-1}}(0, F_{\text{KKT}}(z)) \quad \text{whenever } \|z - z_*\|_P \leq \delta. \quad (7)$$

Here $\mathbf{dist}_{P^{-1}}$ denotes distance in the dual norm induced by P^{-1} .

The preceding discussion is summarized in Figure 2. For the convergence analysis, the important consequence is the final implication: primal–dual nondegeneracy yields strong metric subregularity of the reduced KKT mapping F_{KKT} . This is the error bound used in Section 5. We therefore isolate the corresponding assumption on the limit point of the PDHG sequence.

Assumption 3. *The PDHG iteration (2) converges to a KKT point (X_*, y_*, S_*) satisfying primal–dual nondegeneracy (6).*

4 Local linear convergence under strict complementarity

The analysis in this section assumes the strict-complementarity regime from [Assumption 2](#). This assumption is used in two ways. First, $Z_\star = X_\star - \tau S_\star$ is nonsingular, so the projection onto $\mathbb{S}_+^n$ is smooth near $Z_\star$. Second, the local KKT set has a fixed face structure. These two facts reduce the convergence analysis to a linearized fixed-point map on a smooth local model.

For ease of presentation, we assume without loss of generality that the limit points $X_\star$ and $S_\star$ are diagonal, *i.e.*, $Q_\star = I$ in (4). This assumption does not limit the scope of our conclusions because we can readily construct a pair of SDPs equivalent to (1) and generate PDHG iterates $(\tilde{X}_k, \tilde{y}_k, \tilde{S}_k)$ orthogonally similar to the iterates (X_k, y_k, S_k) generated by (2); see [35] for a detailed construction.

4.1 Local KKT geometry under strict complementarity

Strict complementarity identifies the two minimal faces that contain the nearby primal and dual slack solutions. More precisely, any other KKT point (X, y, S) must satisfy $\langle X, S_\star \rangle = 0$ and $\langle X_\star, S \rangle = 0$. Since $X, S, X_\star, S_\star$ are all positive semidefinite, these equalities force X to lie in the face exposed by $S_\star$, and S to lie in the complementary face exposed by $X_\star$. Thus, locally, the KKT set is a relatively open subset of an affine space obtained by restricting X and $S = C - \mathcal{A}^*y$ to their identified faces. The next lemma records this local representation.

Lemma 1. *Suppose [Assumptions 1](#) and [2](#) hold. There is a neighborhood of $(X_\star, y_\star)$ such that the KKT set is locally represented by*

$$\mathcal{M} = \{(X, y) \mid \mathcal{A}X = b, X \in \mathcal{U}, C - \mathcal{A}^*y \in \mathcal{V}, X_{11} \succ 0, (C - \mathcal{A}^*y)_{22} \succ 0\}, \quad (8)$$

where the linear subspaces $\mathcal{U}$ and $\mathcal{V}$ are defined by

$$\mathcal{U} = \left\{ \begin{bmatrix} U_{11} & 0 \\ 0 & 0 \end{bmatrix} \mid U_{11} \in \mathbb{S}^r \right\}, \quad \mathcal{V} = \left\{ \begin{bmatrix} 0 & 0 \\ 0 & V_{22} \end{bmatrix} \mid V_{22} \in \mathbb{S}^{n-r} \right\}.$$

Equivalently, for all (X, y) sufficiently close to $(X_\star, y_\star)$,

$$(X, y, C - \mathcal{A}^*y) \in \Omega_\star \iff (X, y) \in \mathcal{M}.$$

Moreover, $\mathcal{M}$ is a relatively open subset of an affine space, and its tangent space is

$$\mathcal{T}_\mathcal{M} = \{(U, v) \mid U \in \mathcal{U}, \mathcal{A}U = 0, \mathcal{A}^*v \in \mathcal{V}\}. \quad (9)$$

Proof. Let (X, y, S) be a KKT point near $(X_\star, y_\star, S_\star)$. Since X is primal optimal and $(y_\star, S_\star)$ is dual optimal,

$$\langle C, X \rangle - b^\top y_\star = \langle X, S_\star \rangle = 0.$$

As $X \succeq 0$ and $S_\star \succeq 0$, this implies $XS_\star = 0$. Since $S_\star$ is positive definite on the right bottom block in (4), we obtain $X \in \mathcal{U}$. Similarly, primal optimality of $X_\star$ and dual optimality of (y, S) gives $\langle X_\star, S \rangle = 0$. Since $S \succeq 0$ and $X_\star$ is positive definite on the first block, we obtain $S \in \mathcal{V}$. For all KKT points sufficiently close to $(X_\star, y_\star, S_\star)$, the positive blocks X_{11} and S_{22} remain positive definite. Hence $(X, y) \in \mathcal{M}$.

Conversely, if $(X, y) \in \mathcal{M}$ and $S = C - \mathcal{A}^*y$, then $X \succeq 0$, $S \succeq 0$, $\mathcal{A}X = b$, $\mathcal{A}^*y + S = C$, and $XS = 0$ by the block structure. Thus $(X, y, S) \in \Omega_\star$. Finally, $\mathcal{M}$ is an open subset of the affine space obtained by replacing the two positive-definiteness inequalities in (8) by the linear equalities. Differentiating those affine constraints gives (9). $\square$

4.2 Resolvent representation and local linearization of PDHG

We next rewrite PDHG as a fixed-point iteration in a metric adapted to the primal–dual coupling. This representation has two advantages. First, it gives firm nonexpansiveness of the PDHG map in the natural P -metric. Second, under strict complementarity, it allows us to differentiate the map locally by differentiating the projection onto the positive semidefinite cone.

Define the self-adjoint block operator

$$P = \begin{bmatrix} \tau^{-1}\text{Id} & \mathcal{A}^* \\ \mathcal{A} & \sigma^{-1}I_m \end{bmatrix}. \quad (10)$$

The stepsize condition (3) implies that P is positive definite. Consequently, we define the inner product and norm associated with P :

$$\begin{aligned} \langle (U, v), (W, w) \rangle_P &= \langle U, \tau^{-1}W + \mathcal{A}^*w \rangle + v^\top (\mathcal{A}W + \sigma^{-1}w), \\ \|(U, v)\|_P^2 &= \tau^{-1}\|U\|_F^2 + 2v^\top \mathcal{A}U + \sigma^{-1}\|v\|_2^2. \end{aligned} \quad (11)$$

It is known that PDHG (2) is the preconditioned proximal point method for F_{KKT} in the metric P [32, 34]. Indeed, the update (2) can be written as

$$0 \in F_{\text{KKT}}(z_{k+1}) + P(z_{k+1} - z_k), \quad \text{where } z_k = (X_k, y_k), \quad (12)$$

or equivalently,

$$z_{k+1} = Rz_k, \quad \text{where } R = (P + F_{\text{KKT}})^{-1}P.$$

In other words, R is the P -resolvent of F_{KKT} . In particular, since F_{KKT} is maximal monotone and $P \succ 0$, the PDHG map R is firmly nonexpansive in the P -metric:

$$\|Rz - Rz'\|_P^2 \leq \langle Rz - Rz', z - z' \rangle_P \quad \text{for all } z, z'; \quad (13)$$

see, *e.g.*, [8, Proposition 23.8].

The rest of this subsection studies the local differential properties of R near the local KKT manifold (8). Those properties are built on the directional derivative of the positive-semidefinite cone projector $\Pi_{\mathbb{S}_+^n}$; see [52, Theorem 4.6] and [35, Lemma 1]. More precisely, given a nonsingular matrix $Z \in \mathbb{S}^n$, denote its eigen-decomposition by

$$Z = Q \text{diag}(\lambda_1, \dots, \lambda_n) Q^\top, \quad \text{where } \lambda_1 \geq \dots \geq \lambda_r > 0 > \lambda_{r+1} \geq \dots \geq \lambda_n$$

and $Q \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. Then, the function $\Pi_{\mathbb{S}_+^n}$ is Fréchet differentiable and its Fréchet differential at Z for $H \in \mathbb{S}^n$ is a linear operator $J: \mathbb{S}^n \rightarrow \mathbb{S}^n$ given by

$$JH = D\Pi_{\mathbb{S}_+^n}(Z)[H] = Q(\Gamma \circ (Q^\top H Q))Q^\top,$$

where the symmetric $n \times n$ matrix Γ is defined as

$$\Gamma = \begin{bmatrix} \mathbf{1}_r \mathbf{1}_r^\top & \Theta^\top \\ \Theta & 0 \end{bmatrix}.$$

Here, $\mathbf{1}_r \mathbf{1}_r^\top$ is the all-ones matrix of size $r \times r$ and $\Theta \in \mathbb{R}^{(n-r) \times r}$ captures the off-block-diagonal part in Γ :

$$\Theta_{ij} = \frac{\lambda_j}{\lambda_j - \lambda_{i+r}} \in (0, 1) \quad \text{for } i \in [n-r] \text{ and } j \in [r].$$

Let $\bar{z} = (\bar{X}, \bar{y}) \in \mathcal{M}$ be close to $z_\star = (X_\star, y_\star)$, and set

$$\bar{S} = C - \mathcal{A}^* \bar{y}, \quad \bar{Z} = \bar{X} - \tau \bar{S}, \quad J_{\bar{z}} = D\Pi_{\mathbb{S}_+^n}(\bar{Z}).$$

By the local representation (8),

$$\bar{Z} = \begin{bmatrix} \bar{X}_{11} & 0 \\ 0 & -\tau \bar{S}_{22} \end{bmatrix}, \quad \bar{X}_{11} \succ 0, \quad \bar{S}_{22} \succ 0.$$

Hence $\bar{Z}$ is nonsingular and R is continuously differentiable near $\bar{z}$. Moreover, the derivative of R at $\bar{z}$ is the linear map $G_{\bar{z}} = D R(\bar{z})$ given by

$$G_{\bar{z}} \begin{bmatrix} U \\ v \end{bmatrix} = \begin{bmatrix} J_{\bar{z}}(U + \tau \mathcal{A}^* v) \\ v + \sigma \mathcal{A} U - 2\sigma \mathcal{A} J_{\bar{z}}(U + \tau \mathcal{A}^* v) \end{bmatrix}. \quad (14)$$

In particular, at the limit point (X_*, y_*, S_*) of PDHG, write $Z_* = X_* - \tau S_*$, $J_* = J_{z_*} = D \Pi_{\mathbb{S}_+^n}(Z_*)$, and $G_* = G_{z_*}$.

4.3 Fixed directions and normal contraction

In this subsection, we use the differential description of the PDHG map R to separate the tangent and normal directions to the local KKT set $\mathcal{M}$. We show that the tangent directions $\mathcal{T}_{\mathcal{M}}$ are exactly the set of fixed-points of G_* , and that all directions normal to $\mathcal{M}$ are contracted uniformly.

The first lemma identifies the fixed-point set of the local linearization.

Lemma 2. *Suppose Assumptions 1 and 2 hold. For every $\bar{z} \in \mathcal{M}$ sufficiently close to z_* ,*

$$\text{Fix}(G_{\bar{z}}) = \mathcal{T}_{\mathcal{M}}.$$

In particular, $\text{Fix}(G_) = \mathcal{T}_{\mathcal{M}}$.*

Proof. Let $\bar{z} \in \mathcal{M}$ and (U, v) be a fixed point of $G_{\bar{z}}$; i.e., $G_{\bar{z}}(U, v) = (U, v)$. Set $H = U + \tau \mathcal{A}^* v$. Then the fixed-point equation gives

$$U = J_{\bar{z}} H, \quad \mathcal{A}(U - 2J_{\bar{z}} H) = 0.$$

Substituting $U = J_{\bar{z}} H$ into the second equation obtains $\mathcal{A}U = 0$. Moreover, $\tau \mathcal{A}^* v = H - U = (I - J_{\bar{z}})H$. Taking the inner product with $U = J_{\bar{z}} H$ gives

$$0 = \tau \langle \mathcal{A}U, v \rangle = \langle J_{\bar{z}} H, (I - J_{\bar{z}})H \rangle. \quad (15)$$

Let H be partitioned as

$$H = \begin{bmatrix} H_{11} & H_{21}^\top \\ H_{21} & H_{22} \end{bmatrix}.$$

Then,

$$J_{\bar{z}} H = \begin{bmatrix} H_{11} & (\Theta \circ H_{21})^\top \\ \Theta \circ H_{21} & 0 \end{bmatrix}, \quad (I - J_{\bar{z}})H = \begin{bmatrix} 0 & ((I - \Theta) \circ H_{21})^\top \\ (I - \Theta) \circ H_{21} & H_{22} \end{bmatrix},$$

where the matrix $\Theta \in \mathbb{R}^{(n-r) \times r}$ satisfies $\Theta_{ij} \in (0, 1)$ for all i, j . Hence the zero inner product in (15) forces $H_{21} = 0$. Consequently,

$$U = J_{\bar{z}} H \in \mathcal{U}, \quad \mathcal{A}^* v = \tau^{-1}(I - J_{\bar{z}})H \in \mathcal{V}.$$

Combining with $\mathcal{A}U = 0$ yields $(U, v) \in \mathcal{T}_{\mathcal{M}}$.

Conversely, if $(U, v) \in \mathcal{T}_{\mathcal{M}}$, then $U \in \mathcal{U}$, $\mathcal{A}U = 0$, and $\mathcal{A}^* v \in \mathcal{V}$. Hence $J_{\bar{z}}(U + \tau \mathcal{A}^* v) = U$, and the second component of $G_{\bar{z}}$ in (14) becomes $v + \sigma \mathcal{A}U - 2\sigma \mathcal{A}U = v$. Hence $G_{\bar{z}}(U, v) = (U, v)$. $\square$

The second lemma characterizes the contraction properties of vectors normal to $\mathcal{T}_{\mathcal{M}}$. Let Π^P denote the P -orthogonal projector onto $\mathcal{T}_{\mathcal{M}}$:

$$\Pi^P(z) = \underset{z' \in \mathcal{T}_{\mathcal{M}}}{\operatorname{argmin}} \frac{1}{2} \|z - z'\|_P^2.$$

Lemma 3. Suppose [Assumptions 1](#) and [2](#) hold. There exists $\epsilon > 0$ such that the set $\mathcal{C}_\epsilon = (\mathbf{aff} \mathcal{M}) \cap \mathcal{B}_\epsilon^P(z_\star)$ is a subset of $\mathcal{M}$. Moreover, there exist $\delta > 0$ and $\alpha \in (0, 1)$ such that for every $\bar{z} \in \mathcal{C}_\epsilon$ and every $u \perp_P \mathcal{T}_\mathcal{M}$ with $\|u\|_P \leq \delta$, it holds that

$$R(\bar{z} + u) = \bar{z} + G_{\bar{z}}u + r_{\bar{z}}(u), \quad (16)$$

where

$$\|(I - \Pi^P)G_{\bar{z}}u\|_P \leq \alpha\|u\|_P \quad (17)$$

and

$$\|r_{\bar{z}}(u)\|_P \leq \frac{1 - \alpha}{2}\|u\|_P. \quad (18)$$

Consequently, there is a constant $\rho \in (0, 1)$ such that

$$\mathbf{dist}_P(Rz, \mathcal{M}) \leq \rho \mathbf{dist}_P(z, \mathcal{M}) \quad (19)$$

for all z sufficiently close to $z_\star$.

Proof. Since $\mathcal{M}$ is relatively open in its affine hull and $z_\star \in \mathcal{M}$, there exists $\epsilon > 0$ such that

$$\mathcal{C}_\epsilon := (\mathbf{aff} \mathcal{M}) \cap \mathcal{B}_\epsilon^P(z_\star) \subseteq \mathcal{M}.$$

The set $\mathcal{C}_\epsilon$ is compact, and the map R is continuously differentiable on an open neighborhood of $\mathcal{C}_\epsilon$. Moreover, every $\bar{z} \in \mathcal{C}_\epsilon$ is a KKT point, so $R\bar{z} = \bar{z}$.

Fix $\bar{z} \in \mathcal{C}_\epsilon$. Applying firm nonexpansiveness of R in the P -metric to the pair $\bar{z} + th$ and $\bar{z}$, dividing by t^2 , and letting $t \downarrow 0$, gives

$$\|G_{\bar{z}}h\|_P^2 \leq \langle G_{\bar{z}}h, h \rangle_P \quad \text{for all } h.$$

Consequently,

$$\|G_{\bar{z}}h\|_P^2 \leq \langle G_{\bar{z}}h, h \rangle_P \leq \|G_{\bar{z}}h\|_P \|h\|_P,$$

and hence $\|G_{\bar{z}}h\|_P \leq \|h\|_P$. Suppose equality holds for some nonzero h . Then $G_{\bar{z}}h \neq 0$, and equality must also hold in the second inequality above. Thus $G_{\bar{z}}h$ and h are positively collinear in the P -inner product. Since their P -norms are equal, we must have $G_{\bar{z}}h = h$. By [Lemma 2](#), this implies $h \in \mathcal{T}_\mathcal{M}$.

We now prove that the contraction is uniform in $\bar{z}$. Consider the compact set

$$\mathcal{D}_\epsilon = \{(\bar{z}, u) \mid \bar{z} \in \mathcal{C}_\epsilon, u \perp_P \mathcal{T}_\mathcal{M}, \|u\|_P = 1\}.$$

Then the function $(\bar{z}, u) \mapsto \|(I - \Pi^P)G_{\bar{z}}u\|_P$ is continuous on $\mathcal{D}_\epsilon$. If its maximum were equal to one, then for some $(\bar{z}, u) \in \mathcal{D}_\epsilon$,

$$1 = \|(I - \Pi^P)G_{\bar{z}}u\|_P \leq \|G_{\bar{z}}u\|_P \leq \|u\|_P = 1.$$

Thus $\|G_{\bar{z}}u\|_P = \|u\|_P$. By the equality case above, $u \in \mathcal{T}_\mathcal{M}$, contradicting $u \perp_P \mathcal{T}_\mathcal{M}$ and $\|u\|_P = 1$. Hence the maximum is strictly less than one and can be denoted by $\alpha \in (0, 1)$. By homogeneity, this gives [\(17\)](#) for all $u \perp_P \mathcal{T}_\mathcal{M}$.

It remains to make the linear approximation uniform. Since R is continuously differentiable on a neighborhood of the compact set $\mathcal{C}_\epsilon$, its derivative is uniformly continuous there. Hence there exists $\delta > 0$ such that, for all $\bar{z} \in \mathcal{C}_\epsilon$ and all $u \perp_P \mathcal{T}_\mathcal{M}$ with $\|u\|_P \leq \delta$, the residual $r_{\bar{z}}(u) = R(\bar{z} + u) - \bar{z} - G_{\bar{z}}u$ satisfies

$$\|r_{\bar{z}}(u)\|_P \leq \frac{1 - \alpha}{2}\|u\|_P.$$

This proves [\(16\)](#) and [\(18\)](#).

We now prove the distance contraction. Choose a neighborhood $\mathcal{O}$ of $z_\star$ such that, for every $z \in \mathcal{O}$, the P -orthogonal projection $\bar{z} := \Pi_{\mathbf{aff} \mathcal{M}}^P(z)$ belongs to $\mathcal{C}_\epsilon$ and the normal component $u := z - \bar{z}$ satisfies $\|u\|_P \leq \delta$. Since $\mathbf{aff} \mathcal{M}$ is affine, $u \perp_P \mathcal{T}_\mathcal{M}$. Since also $\bar{z} \in \mathcal{M}$, we have

$$\mathbf{dist}_P(z, \mathcal{M}) = \mathbf{dist}_P(z, \mathbf{aff} \mathcal{M}) = \|u\|_P.$$

The first equality follows because $\mathcal{M}$ agrees locally with its affine hull around $\bar{z}$.

Using the expansion above, $Rz = \bar{z} + G_{\bar{z}}u + r_{\bar{z}}(u)$. The vector $\Pi^P(G_{\bar{z}}u + r_{\bar{z}}(u))$ belongs to $\mathcal{T}_{\mathcal{M}}$. Hence,

$$\hat{z} := \bar{z} + \Pi^P(G_{\bar{z}}u + r_{\bar{z}}(u)) \in \text{aff } \mathcal{M}.$$

By shrinking the neighborhood $\mathcal{O}$ if necessary, we may ensure that $\hat{z} \in \mathcal{M}$, since $\mathcal{M}$ is relatively open in $\text{aff } \mathcal{M}$ and $\bar{z} \in \mathcal{C}_\epsilon \subseteq \mathcal{M}$. Therefore,

$$\begin{aligned} \text{dist}_P(Rz, \mathcal{M}) &\leq \|Rz - (\bar{z} + \Pi^P(G_{\bar{z}}u + r_{\bar{z}}(u)))\|_P \\ &= \|(I - \Pi^P)(G_{\bar{z}}u + r_{\bar{z}}(u))\|_P \\ &\leq \|(I - \Pi^P)G_{\bar{z}}u\|_P + \|(I - \Pi^P)r_{\bar{z}}(u)\|_P \\ &\leq \alpha\|u\|_P + \frac{1-\alpha}{2}\|u\|_P \\ &= \frac{1+\alpha}{2} \text{dist}_P(z, \mathcal{M}). \end{aligned}$$

Thus (19) holds on $\mathcal{O}$ with $\rho = \frac{1+\alpha}{2} < 1$. □

4.4 Local linear convergence

We now combine the preceding ingredients. The local KKT set is a smooth affine manifold in the relevant neighborhood, the derivative of the PDHG map fixes precisely the tangent directions to this manifold, and the normal component is uniformly contracted. Since the iterates converge to the strictly complementary limit point, they eventually enter this neighborhood, and the normal-contraction estimate can be iterated.

Theorem 1 (Local R -linear convergence under strict complementarity). *Suppose Assumptions 1 and 2 hold and the PDHG stepsizes τ, σ satisfy (3). Let $\{(X_k, y_k)\}$ be the sequence generated by PDHG (2), and set $S_k = C - \mathcal{A}^*y_k$. Let $z_k = (X_k, y_k)$ and $z_\star = (X_\star, y_\star)$, and let $\mathcal{M}$ be the local KKT manifold defined in (8). Then there exist $\rho_{\text{SC}} \in (0, 1)$ and an integer $k_{\text{SC}} \in \mathbb{N}$ such that for all $k \geq k_{\text{SC}}$,*

$$\text{dist}_P(z_{k+1}, \mathcal{M}) \leq \rho_{\text{SC}} \text{dist}_P(z_k, \mathcal{M}). \quad (20)$$

Moreover, there exists $\kappa_{\text{SC}} > 0$ such that for all $k \geq k_{\text{SC}}$,

$$\text{dist}_P(z_k, \mathcal{M}) \leq \kappa_{\text{SC}} \rho_{\text{SC}}^{k-k_{\text{SC}}}, \quad (21)$$

and

$$\text{dist}((X_k, y_k, S_k), \Omega_\star) \leq \kappa_{\text{SC}} \rho_{\text{SC}}^{k-k_{\text{SC}}}. \quad (22)$$

Finally, it holds for all $k \geq k_{\text{SC}}$ that

$$\|X_{k+1} - X_k\|_F + \|y_{k+1} - y_k\|_2 + \|S_{k+1} - S_k\|_F \leq \kappa_{\text{SC}} \rho_{\text{SC}}^{k-k_{\text{SC}}}, \quad (23)$$

$$\|X_k - X_\star\|_F + \|y_k - y_\star\|_2 + \|S_k - S_\star\|_F \leq \kappa_{\text{SC}} \rho_{\text{SC}}^{k-k_{\text{SC}}}. \quad (24)$$

If additionally $(X_\star, y_\star, S_\star)$ is locally isolated in $\Omega_\star$, then z_k converges Q -linearly to $z_\star$ in the P -metric; equivalently, (X_k, y_k, S_k) converges Q -linearly to $(X_\star, y_\star, S_\star)$.

Proof. Since z_k converges to $z_\star$, Lemma 3 applies for sufficiently large k . Thus, there exists $k_{\text{SC}} \in \mathbb{N}$ such that

$$\text{dist}_P(z_{k+1}, \mathcal{M}) \leq \rho_{\text{SC}} \text{dist}_P(z_k, \mathcal{M}) \quad \text{for } k \geq k_{\text{SC}},$$

for some $\rho_{\text{SC}} \in (0, 1)$. Iteration gives (21), with a suitable choice of the constant κ_1 .

We next estimate the steps. Let $\bar{z}_k = \Pi_{\text{aff } \mathcal{M}}^P(z_k)$. For $k \geq k_{\text{SC}}$, the local construction in Lemma 3 gives

$$\bar{z}_k \in \mathcal{M}, \quad \|z_k - \bar{z}_k\|_P = \text{dist}_P(z_k, \mathcal{M}).$$

Since $\bar{z}_k \in \mathcal{M}$, it is a fixed point of R . Therefore, the nonexpansiveness of R in the P -metric yields

$$\begin{aligned}\|z_{k+1} - z_k\|_P &= \|Rz_k - z_k\|_P \\ &\leq \|Rz_k - R\bar{z}_k\|_P + \|z_k - \bar{z}_k\|_P \\ &\leq 2\|z_k - \bar{z}_k\|_P = 2\mathbf{dist}_P(z_k, \mathcal{M}).\end{aligned}$$

Together with (21), this gives $\|z_{k+1} - z_k\|_P \leq \kappa_1 \rho_{\text{SC}}^{k-k_{\text{SC}}}$ for some $\kappa_1 \geq 0$. Since P is positive definite, the P -norm is equivalent to the product norm on $\mathbb{S}^n \times \mathbb{R}^m$. Together with $S_{k+1} - S_k = -\mathcal{A}^*(y_{k+1} - y_k)$, this gives (23).

We now pass from $\mathcal{M}$ to the KKT set. Write $\bar{z}_k = (\bar{X}_k, \bar{y}_k)$ and $\bar{S}_k = C - \mathcal{A}^* \bar{y}_k$. Since $\bar{z}_k \in \mathcal{M}$, we have $(\bar{X}_k, \bar{y}_k, \bar{S}_k) \in \Omega_\star$. Hence, for a constant $\kappa_2 > 0$,

$$\begin{aligned}\mathbf{dist}((X_k, y_k, S_k), \Omega_\star) &\leq \|X_k - \bar{X}_k\|_F + \|y_k - \bar{y}_k\|_2 + \|S_k - \bar{S}_k\|_F \\ &\leq \kappa_2 \|z_k - \bar{z}_k\|_P = \kappa_2 \mathbf{dist}_P(z_k, \mathcal{M}).\end{aligned}$$

Combining this bound with (21) proves (22).

Finally, summing the tail of (23) gives

$$\begin{aligned}\|X_k - X_\star\|_F + \|y_k - y_\star\|_2 + \|S_k - S_\star\|_F &\leq \sum_{j=k}^{\infty} (\|X_{j+1} - X_j\|_F + \|y_{j+1} - y_j\|_2 + \|S_{j+1} - S_j\|_F) \\ &\leq \kappa_3 \sum_{j=k}^{\infty} \rho_{\text{SC}}^{j-k_{\text{SC}}} \leq \frac{\kappa_3}{1 - \rho_{\text{SC}}} \rho_{\text{SC}}^{k-k_{\text{SC}}},\end{aligned}$$

for some constant $\kappa_3 > 0$. This proves (24). (The constant κ_{SC} in the theorem statement can be chosen as $\kappa_{\text{SC}} = \max\{\kappa_1, \kappa_2, \kappa_3/(1 - \rho_{\text{SC}})\}$.)

If $(X_\star, y_\star, S_\star)$ is locally isolated in $\Omega_\star$, then $\mathcal{M} = \{z_\star\}$ locally. For all sufficiently large k , $\mathbf{dist}_P(z_k, \mathcal{M}) = \|z_k - z_\star\|_P$ and (20) becomes

$$\|z_{k+1} - z_\star\|_P \leq \rho_{\text{SC}} \|z_k - z_\star\|_P.$$

Thus z_k converges Q -linearly to $z_\star$. Since $S_k - S_\star = -\mathcal{A}^*(y_k - y_\star)$, the same conclusion holds for (X_k, y_k, S_k) . $\square$

5 Local linear convergence under primal–dual nondegeneracy

The previous section uses strict complementarity to obtain a smooth local representation of the PSD-cone projection and then analyzes the derivative of the PDHG fixed-point map. We now consider a different regularity regime. When the limit KKT point is primal–dual nondegenerate, the KKT mapping is strongly metrically subregular at the solution. This error bound bypasses the need to differentiate the PSD-cone projection: the usual proximal-point contraction argument, applied to the P -resolvent representation of PDHG, directly gives local Q -linear convergence to the KKT point.

Theorem 2 (Local Q -linear convergence under primal–dual nondegeneracy). *Suppose Assumptions 1 and 3 hold and the PDHG stepsizes τ, σ satisfy (3). Then there exist $\rho_{\text{ND}} \in (0, 1)$, $\kappa_{\text{ND}} > 0$, and an integer k_{ND} such that, for all $k \geq k_{\text{ND}}$,*

$$\|(X_{k+1}, y_{k+1}) - (X_\star, y_\star)\|_P \leq \rho_{\text{ND}} \|(X_k, y_k) - (X_\star, y_\star)\|_P,$$

and hence

$$\|X_k - X_\star\|_F + \|y_k - y_\star\|_2 + \|S_k - S_\star\|_F \leq \kappa_{\text{ND}} \rho_{\text{ND}}^{k-k_{\text{ND}}}.$$

Proof. Let $(X_\star, y_\star, S_\star)$ be the limit point of PDHG. From [Section 3.4](#), primal–dual nondegeneracy implies that the KKT point $(X_\star, y_\star, S_\star)$ is unique and that the KKT mapping F_{KKT} is strongly metrically subregular at $((X_\star, y_\star), 0)$; i.e., [\(7\)](#) holds.

Let $z_+ = Rz$. The inclusion [\(12\)](#) implies that $P(z - z_+) \in F_{\text{KKT}}(z_+)$. This means that $P(z - z_+)$ is an admissible residual of the KKT mapping at z_+ :

$$\mathbf{dist}_{P^{-1}}(0, F_{\text{KKT}}(z_+)) \leq \|P(z - z_+)\|_{P^{-1}}.$$

Since R is nonexpansive in the P -norm and $Rz_\star = z_\star$, there is a neighborhood $\mathcal{O}$ of $z_\star$ such that Rz belongs to the neighborhood in which [\(7\)](#) holds whenever $z \in \mathcal{O}$. Hence, for all $z \in \mathcal{O}$,

$$\|z_+ - z_\star\|_P \leq \mu \|P(z - z_+)\|_{P^{-1}} = \mu \|z - z_+\|_P. \quad (25)$$

On the other hand, firm nonexpansiveness of R and $z_\star \in \text{Fix}(R)$ give

$$\langle z_+ - z_\star, z - z_+ \rangle_P \geq 0.$$

Expanding $z - z_\star = (z_+ - z_\star) + (z - z_+)$ yields

$$\|z_+ - z_\star\|_P^2 + \|z - z_+\|_P^2 \leq \|z - z_\star\|_P^2. \quad (26)$$

Combining [\(25\)](#) and [\(26\)](#) yields

$$(1 + \mu^{-2}) \|z_+ - z_\star\|_P^2 \leq \|z - z_\star\|_P^2.$$

Thus, with $\rho_{\text{ND}} = \mu / \sqrt{1 + \mu^2} < 1$,

$$\|Rz - z_\star\|_P \leq \rho_{\text{ND}} \|z - z_\star\|_P$$

for all z sufficiently close to $z_\star$. Since $z_k \rightarrow z_\star$, the estimate applies for all sufficiently large k and gives the asserted Q -linear convergence. The estimate for S_k follows from $S_k - S_\star = -\mathcal{A}^*(y_k - y_\star)$ and the equivalence of norms. $\square$

6 Discussion

[Sections 4](#) and [5](#) establish local linear convergence of PDHG for SDP under two different regularity assumptions: strict complementarity and primal–dual nondegeneracy. We now discuss the scope of these results and the limitations of the arguments used to prove them. We begin with a simple SDP instance showing that, without such regularity, PDHG may converge only sublinearly. We then compare the proof mechanism with the polyhedral behavior of PDHG for LP and with related ADMM analyses for SDP. We also show that PDHG exhibits a finite-time rank identification phenomenon, parallel to basis identification for LP and related identification properties of ADMM for SDP. Finally, we close with several open questions.

6.1 A sublinear example without regularity

From [Theorems 1](#) and [2](#), local linear convergence follows from regularity conditions such as strict complementarity or primal–dual nondegeneracy. The next example suggests that some regularity is in fact needed. It gives a three-dimensional SDP for which PDHG converges to a unique KKT point, but the limit point is degenerate and not strictly complementary, and the convergence along a slow branch is only sublinear. The mechanism is the nonlinear behavior of the PSD-cone projection at a singular matrix.

Example 1 (Sublinear convergence). Consider the following primal SDP

$$\begin{aligned} & \text{minimize} && X_{33} \\ & \text{subject to} && X_{11} = 1, \quad X_{22} + (X_{13} + X_{31})/2 = 0 \\ & && X \in \mathbb{S}_+^3. \end{aligned}$$

The unique solution is $X_\star = E_{11}$, $y_\star = 0$, and $S_\star = E_{33}$, which fails strict complementarity. The point also fails dual nondegeneracy, because $E_{12} \in \text{Null}(\mathcal{A}) \cap \mathcal{T}_{S_\star}^\perp$.

Take $\tau = \sigma = 1/2$. The purpose of the following calculation is to give a formal local asymptotic reduction of the PDHG map near the degenerate KKT point and identify the leading slow dynamics. We seek a local one-dimensional expansion, parametrized by $\alpha = X_{12}$, of the form

$$X(\alpha) = \begin{bmatrix} 1 + O(\alpha^4) & \alpha & -\alpha^2 + O(\alpha^4) \\ \alpha & \alpha^2 + O(\alpha^4) & -\alpha^3 + O(\alpha^5) \\ -\alpha^2 + O(\alpha^4) & -\alpha^3 + O(\alpha^5) & \alpha^4 + O(\alpha^6) \end{bmatrix}, \quad y(\alpha) = (-3\alpha^4 + O(\alpha^6), -2\alpha^2 + O(\alpha^4)).$$

Substitution of these expansions into one PDHG step gives the leading-order behavior of the reduced coordinate. Indeed, let $Z(\alpha) = X(\alpha) - \tau(C - \mathcal{A}^*y(\alpha))$. Then,

$$Z(\alpha) = \begin{bmatrix} 1 + O(\alpha^4) & \alpha & -\frac{3}{2}\alpha^2 + O(\alpha^4) \\ \alpha & O(\alpha^4) & -\alpha^3 + O(\alpha^5) \\ -\frac{3}{2}\alpha^2 + O(\alpha^4) & -\alpha^3 + O(\alpha^5) & -\frac{1}{2} + O(\alpha^4) \end{bmatrix}.$$

For sufficiently small $\alpha > 0$, the matrix $Z(\alpha)$ has exactly one positive eigenvalue. If the corresponding eigenvector is normalized to have first component one, then

$$v(\alpha) = \begin{bmatrix} 1 \\ \alpha - \alpha^3 + O(\alpha^5) \\ -\alpha^2 + O(\alpha^4) \end{bmatrix}, \quad \lambda_+(\alpha) = 1 + \alpha^2 + O(\alpha^4).$$

Direct calculation then shows that

$$\Pi_{\mathbb{S}_+^3}(Z(\alpha)) = \lambda_+(\alpha) \frac{v(\alpha)v(\alpha)^\top}{\|v(\alpha)\|^2} = X(\alpha_+), \quad \text{where } \alpha_+ = (\Pi_{\mathbb{S}_+^3}(Z(\alpha)))_{12} = \alpha - \alpha^3 + O(\alpha^5).$$

The dual update gives the same reduced recursion to the displayed order. Thus the formal reduced dynamics along the slow degenerate direction is

$$\alpha_+ = \alpha - \alpha^3 + O(\alpha^5).$$

This scalar recursion predicts sublinear decay. Indeed, if a positive sequence satisfies

$$\alpha_{k+1} = \alpha_k - \alpha_k^3 + O(\alpha_k^5), \quad \alpha_k \downarrow 0,$$

then

$$\frac{1}{\alpha_{k+1}^2} = \frac{1}{\alpha_k^2} + 2 + O(\alpha_k^2).$$

Equivalently, the leading-order asymptotics are $\alpha_k \sim \frac{1}{\sqrt{2k}}$. Moreover, along the above formal expansion,

$$\|X(\alpha) - X_\star\|_F^2 = 2\alpha^2 + O(\alpha^4), \quad \|y(\alpha) - y_\star\|_2 = O(\alpha^2).$$

Hence the formal model predicts

$$\|X_k - X_\star\|_F = \Theta(k^{-1/2})$$

along the slow direction.

The above calculation provides a local asymptotic explanation of the slow dynamics; the numerical evidence in [Figure 3](#) is consistent with the formal prediction: it shows $k^{-1/2}$ decay of the primal error and stabilization of the scaled quantity $\sqrt{k}\|X_k - X_\star\|_F$.

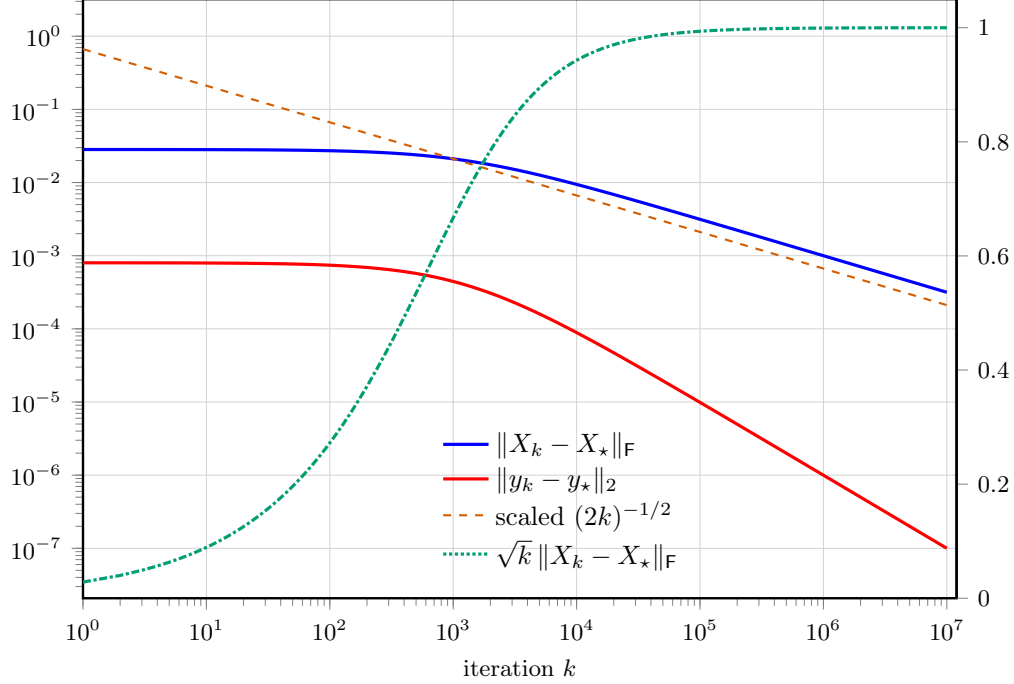


Figure 3: Numerical behavior of PDHG for the SDP in [Example 1](#). The left vertical axis uses a logarithmic scale and shows the distances to the KKT point; the dashed curve is a scaled multiple of $(2k)^{-1/2}$. The right vertical axis uses a linear scale and shows the diagnostic $\sqrt{k} \|X_k - X_*\|_F$, whose stabilization is consistent with $\|X_k - X_*\|_F \asymp k^{-1/2}$.

6.2 Comparison with related linear convergence results

[Example 1](#) illustrates a basic difference between linear programming and semidefinite programming. For LP, the cone is polyhedral and the projection onto the nonnegative orthant is piecewise affine. Once the active set has been identified, PDHG reduces to a linear fixed-point iteration on a fixed subspace. Even before this identification occurs, Hoffman-type error bounds provide a global polyhedral error-bound mechanism. This mechanism, however, is not available for SDP. At a singular matrix, the projection onto the PSD cone has a genuinely nonlinear zero-eigenvalue block. Strict complementarity bypasses this obstruction by making $X_* - \tau S_*$ nonsingular, so that the PSD-cone projection is differentiable in a neighborhood of the point at which it is evaluated. Primal–dual nondegeneracy removes the obstruction in a different way: it gives a local error bound through metric subregularity of the KKT mapping.

The comparison with ADMM is more delicate. Recent analyses of ADMM for SDP [\[35\]](#) study the Douglas–Rachford form of ADMM in the signed matrix variable $Z = X - \sigma S$. The local linearization then acts on Z , and the fixed space of the corresponding linear map describes the directions that are not contracted by the first-order model. In the notation of those analyses, a direction belongs to this fixed space precisely when its off-diagonal block vanishes and its two diagonal blocks satisfy one condition in $\text{Null}(\mathcal{A})$ and one condition in $\text{Range}(\mathcal{A}^*)$. This characterization plays the same structural role as [Lemma 2](#): both identify the neutral directions of the local linearization. Under primal–dual nondegeneracy this fixed space is trivial. Without nondegeneracy, one cannot expect contraction in every direction, and the analysis must instead prove contraction only in the component normal to the local solution set.

There are, however, important differences between the two arguments. The ADMM linearization is expressed in the single matrix variable Z and uses the Frobenius metric together with the orthogonal projector onto $\text{Range}(\mathcal{A}^*)$. The PDHG linearization acts on the pair (X, y) . In the Euclidean metric it is generally non-self-adjoint, and its natural nonexpansiveness property is expressed in the block metric [\(10\)](#). Consequently,

the PDHG proof contracts the P -orthogonal component normal to the local KKT manifold, rather than an ordinary Frobenius-orthogonal component in the Z -variable. Thus the fixed-space calculations in the ADMM and PDHG analyses are analogous, but the residual estimates and the underlying metrics are method-specific.

6.3 Rank identification

PDHG for linear programming, viewed as a special case of SDP, exhibits a two-stage behavior [44]. In the first stage, the method identifies the optimal basis in finitely many iterations; this stage may be sublinear. After identification, the iterates remain on a fixed polyhedral stratum, and the second stage is linear, with a rate governed by the local sharpness constant. For SDP, the analogous issue is whether PDHG identifies, in finite time, the *rank* of the solution to which it converges.

In this subsection we give a partial answer. We show that, once the PDHG iterates are sufficiently close to a strictly complementary limit point, the rank of the primal variable is identified in finitely many iterations and then remains constant. Equivalently, if $X_k \rightarrow X_*$, then $\text{rank } X_k = \text{rank } X_*$ for all sufficiently large k . This finite identification property also follows from the general theory of partial smoothness [22, 39, 63], since the relative interiors of faces of the PSD cone form the relevant active manifolds. Here we give a direct proof for SDP, based only on the spectral form of the projection onto the PSD cone.

Proposition 1 (Finite-time rank identification). *Suppose [Assumptions 1](#) and [2](#) hold. Let $r = \text{rank } X_*$, and define*

$$Z_k = X_k - \tau(C - \mathcal{A}^* y_k) = X_k - \tau S_k, \quad \text{and} \quad \hat{S}_{k+1} = \tau^{-1}(X_{k+1} - Z_k).$$

Then $X_{k+1} = \Pi_{\mathbb{S}_+^n}(Z_k)$, and there exists an integer k_{ID} such that, for all $k \geq k_{\text{ID}}$,

$$\text{inertia}(Z_k) = (r, n-r, 0), \quad \text{rank}(X_{k+1}) = r, \quad \text{rank}(\hat{S}_{k+1}) = n-r, \quad \hat{S}_{k+1} \succeq 0, \quad \langle X_{k+1}, \hat{S}_{k+1} \rangle = 0.$$

Proof. Since (X_k, y_k, S_k) converges to (X_*, y_*, S_*) , it holds that $Z_k \rightarrow Z_*$. The eigenvalues of Z_* consist of the positive eigenvalues of X_* and the negative eigenvalues $-\tau\lambda_i(S_*)$. Let

$$\gamma = \min\{\lambda_{\min}(X_{*,11}), \tau\lambda_{\min}(S_{*,22})\} > 0$$

in the block representation (4). For sufficiently large k , $\|Z_k - Z_*\|_2 < \gamma/2$. Weyl's inequality then gives $\text{inertia}(Z_k) = (r, n-r, 0)$. Since the PSD projection keeps exactly the positive spectral part, $\text{rank}(X_{k+1}) = r$.

If $Z_k = Q \text{diag}(\lambda_1, \dots, \lambda_n) Q^T$ is the eigen-decomposition, then

$$X_{k+1} = Q \text{diag}(\lambda_1^+, \dots, \lambda_n^+) Q^T, \quad \hat{S}_{k+1} = \tau^{-1} Q \text{diag}((-\lambda_1)^+, \dots, (-\lambda_n)^+) Q^T,$$

where $t^+ = \max\{t, 0\}$. The desired properties follow immediately. $\square$

The proof only uses the spectral gap $\lambda_r(Z_*) > 0 > \lambda_{r+1}(Z_*)$, where $Z_* = X_* - \tau S_*$. For complementary $X_*, S_* \succeq 0$, this condition is equivalent to strict complementarity. [Proposition 1](#) shows that the primal sequence $\{X_k\}$ identifies the rank $r = \text{rank } X_*$ after finitely many iterations. However, the dual slack iterates $S_k = C - \mathcal{A}^* y_k$ may not have the identification property: they need not be positive semidefinite, and indeed they may fail to be so infinitely often. [Example 2](#) gives such an example.

Example 2. Consider the SDP

$$\begin{aligned} & \text{minimize} && 0 \\ & \text{subject to} && X_{11} = 1, \quad X_{12} = 0, \quad X_{22} = 1 \\ & && X \in \mathbb{S}_+^2. \end{aligned}$$

The unique KKT point is $X_* = I_2$, $y_* = 0$, $S_* = 0$, which is strictly complementary. Take $\tau = 1$, $\sigma = 3/4$ and initialize at

$$X_0 = \begin{bmatrix} 1 - \epsilon/\sqrt{3} & 0 \\ 0 & 1 \end{bmatrix}, \quad y_0 = 0.$$

where $\epsilon > 0$ is small. The iterates remain in the interior of $\mathbb{S}_+^2$ and can be written concisely as

$$X_k = \begin{bmatrix} 1 + \alpha_k & 0 \\ 0 & 1 \end{bmatrix}, \quad y_k = \begin{bmatrix} \beta_k \\ 0 \\ 0 \end{bmatrix},$$

where the two sequences $\{\alpha_k\}$ and $\{\beta_k\}$ satisfy $\alpha_{k+1} = \alpha_k + \beta_k$ and $\beta_{k+1} = -\frac{3}{4}\alpha_k - \frac{1}{2}\beta_k$ with $\alpha_0 = -\epsilon/\sqrt{3}$ and $\beta_0 = 0$. Eliminating α_k gives

$$\beta_0 = 0, \quad \beta_1 = \epsilon \frac{\sqrt{3}}{4}, \quad \beta_{k+2} - \frac{1}{2}\beta_{k+1} + \frac{1}{4}\beta_k = 0 \text{ for } k \in \mathbb{N}.$$

Direct calculation shows that

$$\beta_k = \epsilon 2^{-k} \sin \frac{k\pi}{3}.$$

Thus, $S_k = C - \mathcal{A}^* y_k = -\beta_k E_{11}$. So $S_k = 0$ for $k = 0, 3, 6, \dots$, while $S_k \neq 0$ for infinitely many other k . Therefore, the rank of S_k never stabilizes to $\text{rank } S_\star = 0$.

From [Theorem 1](#) and [Proposition 1](#), both local linear convergence and finite rank identification hold for PDHG under strict complementarity. The arguments, however, do not establish a causal relation between the two properties. The rank identification of X_k and $\hat{S}_k$ follows from convergence to a strictly complementary solution; no rate estimate is used in this argument.

Conversely, rank identification alone does not imply local linear convergence. After the rank has been identified, one still needs a contraction estimate, an error bound, or a growth condition to control the remaining smooth dynamics. Thus, in contrast with the polyhedral case of PDHG for LP, the relation between identification and the onset of a linear rate is less transparent for SDP. The thresholds are governed by different quantities: rank identification depends on the spectral gap of $Z_\star$, while local linear convergence also depends on the contraction of the linearized PDHG map in directions normal to the local KKT set.

6.4 Open questions

Several questions remain open. First, our proof uses the stepsize condition $\tau\sigma\|\mathcal{A}\|^2 < 1$ and the fully extrapolated update $2X_{k+1} - X_k$. However, the relaxed PDHG iteration

$$\begin{aligned} X_{k+1} &= \Pi_{\mathbb{S}_+^n}(X_k - \tau(C - \mathcal{A}^* y_k)) \\ y_{k+1} &= y_k + \sigma(b - \mathcal{A}(X_{k+1} + \theta(X_{k+1} - X_k))) \end{aligned}$$

is known to converge for $\theta > 1/2$ and $\tau\sigma\|\mathcal{A}\|^2 < 4/(1+2\theta)$ [7]; see also [56] for a different admissible range. It is unclear whether the present argument extends to these larger parameter ranges. In particular, the present proof relies heavily on the proximal point method interpretation of PDHG, whereas the analysis of relaxed PDHG uses a different Lyapunov function.

Second, the available analyses of ADMM [35, 36] and PDHG for SDP appear to use different mechanisms. Both methods are primal–dual proximal methods, and both analyses ultimately depend on the local geometry of the PSD-cone projection. Nevertheless, the ADMM arguments are typically expressed through the signed matrix variable associated with the Douglas–Rachford operator, while the PDHG argument in this paper uses the P -metric resolvent. It would be useful to know whether this distinction is specific to ADMM and PDHG, or whether it reflects a more general classification of primal–dual proximal methods (see, *e.g.*, [17, 31, 37, 45]).

Third, strict complementarity and primal–dual nondegeneracy both lead to local linear convergence, but through different mechanisms. At present, it is not clear whether either condition captures the exact regularity needed for linear convergence, or whether there is a sharper necessary-and-sufficient condition intrinsic to the local KKT geometry and the PDHG fixed-point map. Both assumptions are also a posteriori: they are imposed at the limit point selected by the algorithm, rather than at an arbitrary optimal solution. This is natural, since the local rate is determined by the geometry near the point to which the iterates converge. Nevertheless, it leaves open a more global question: how do primal–dual first-order methods for SDP select a particular KKT point, and how is this selection affected by the curvature and geometry of the positive semidefinite cone?

Table 1: Basic information on the tested SDP instances.

Instance	n	m	τ	σ	Instance	n	m	τ	σ
maxG*	800	800	1.00	1.00	XM-48	144	241	0.85	0.85
QS-20	231	16402	0.40	0.40	XM-93	279	466	0.85	0.85
QS-30	496	77377	0.12	1.20	XM-149	447	746	0.70	0.70
QS-40	861	236202	0.10	1.00	rose13	105	2379	0.30	0.30
pdnd-nosc-100	100	136	0.50	0.50	cnhi110	220	5005	0.35	0.35
pdnd-nosc-200	200	210	0.50	0.50	BQP-r1-20-1	231	20601	0.30	0.30
pdnd-nosc-500	500	378	0.50	0.50					

7 Numerical experiments

In this section, we present numerical evidence supporting the theoretical findings of the paper. The experiments are designed to illustrate the following two phenomena.

1. Local (R -)linear convergence is observed under different combinations of primal–dual nondegeneracy (ND) and strict complementarity (SC).
2. When SC is close to failure, PDHG applied to SDP may converge extremely slowly, and no clear linear regime may be observed within the prescribed computational budget.

All numerical experiments were conducted on a Mac mini equipped with an Apple M4 Pro chip and 48 GB of unified memory, running macOS 15.7.4. The algorithms were implemented in Python. The code and synthetic SDP data used in the experiments are available at

<https://github.com/NumOptLab/pdhg-sdp-linear-conv>.

For the primal–dual pair of SDPs in (1), we define primal infeasibility ϵ_p , dual infeasibility ϵ_d , complementarity residual ϵ_c , and relative objective gap ϵ_{gap} by

$$\epsilon_p = \frac{\|\mathcal{A}X - b\|_2}{1 + \|b\|_2}, \quad \epsilon_d = \max \left\{ 0, \frac{-\lambda_{\min}(S)}{1 + \|C\|_F} \right\}, \quad \epsilon_c = \frac{\langle X, \Pi_{\mathbb{S}_+^n}(S) \rangle}{1 + |\langle C, X \rangle| + |\langle b, y \rangle|}, \quad \epsilon_{\text{gap}} = \frac{|\langle C, X \rangle - b^T y|}{1 + |\langle C, X \rangle| + |\langle b, y \rangle|}.$$

The overall KKT residual is then defined as $\epsilon_{\text{KKT}} := \max\{\epsilon_p, \epsilon_d, \epsilon_c, \epsilon_{\text{gap}}\}$. Unless otherwise specified, the algorithm is terminated when $\epsilon_{\text{KKT}} \leq 10^{-8}$ or when the iteration count reaches 10^5 . The SDP instances used in the experiments are summarized in Table 1. Strict complementarity is assessed numerically by computing the smallest absolute eigenvalue of the final Z -iterate.

7.1 Demonstration of local linear convergence

We first report SDP instances for which some regularity conditions in our analysis appear to hold numerically. The test set contains examples from several application domains, together with synthetic instances constructed to separate the roles of strict complementarity and primal–dual nondegeneracy. In all these experiments, PDHG exhibits a clear local linear convergence regime after a transient phase. We also observe finite identification of the numerical rank of the primal iterate.

MaxCut SDP: SC holds and primal–dual ND holds. Figure 4 reports three representative MaxCut SDP instances. For all three cases, strict complementarity and primal–dual nondegeneracy appear to hold numerically. Consistent with the theory, PDHG enters a local linear convergence regime after a short transient phase.

The three panels show the evolution of the KKT residual, the consecutive fixed-point difference $\|Z_{k+1} - Z_k\|_F$, and the rank of X_k . In each case, the residual and the fixed-point difference eventually decrease at an

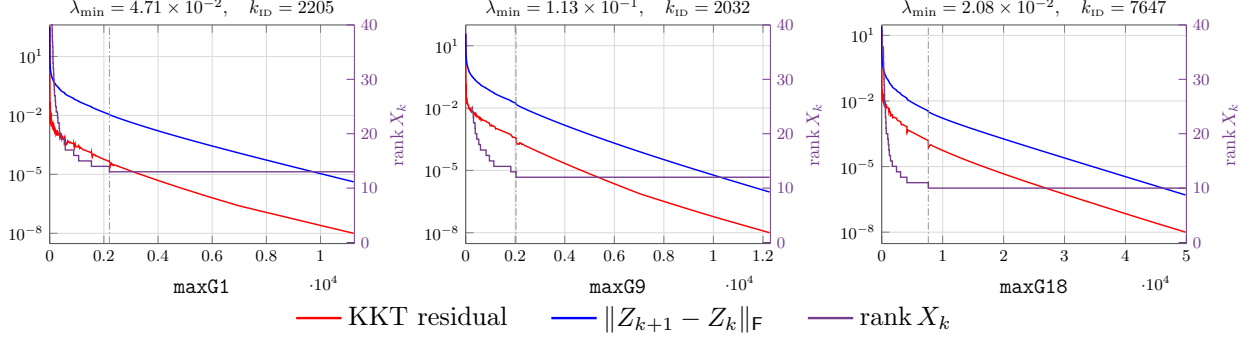


Figure 4: PDHG convergence on MaxCut SDP instances. The red and blue curves report the KKT residual and the fixed-point difference, respectively, on a logarithmic scale; the purple curve shows $\text{rank } X_k$ on the right axis in the standard scale. The dash-dotted vertical line marks the rank-identification iteration k_{ID} , and $\lambda_{\min}$ denotes the smallest absolute eigenvalue of the final Z -iterate.

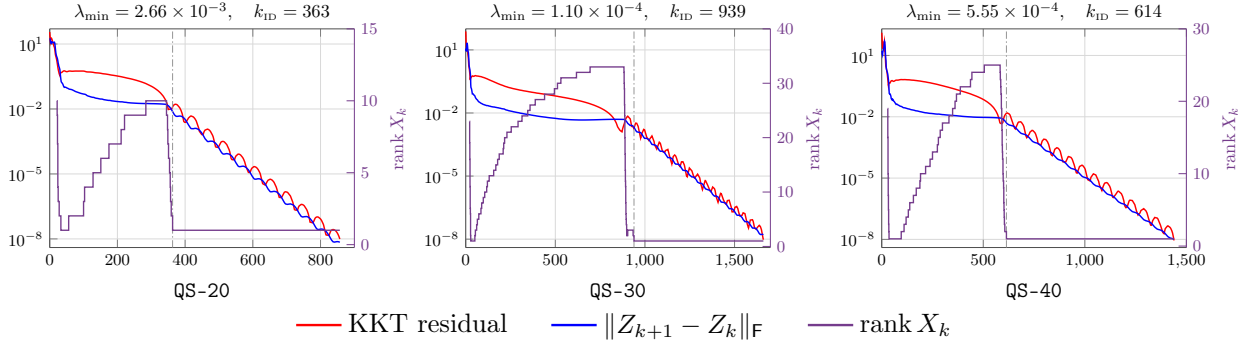


Figure 5: PDHG convergence on random QS problems. The red and blue curves report the KKT residual and the fixed-point difference, respectively, on a logarithmic scale; the purple curve shows $\text{rank } X_k$ on the right axis in the standard scale. The dash-dotted vertical line marks the rank-identification iteration k_{ID} , and $\lambda_{\min}$ denotes the smallest absolute eigenvalue of the final Z -iterate.

approximately linear rate on the semilogarithmic scale. The rank of X_k also stabilizes after finitely many iterations, indicating identification of the numerical rank of the primal solution. The reported values of $\lambda_{\min}$ are bounded away from zero, which is consistent with strict complementarity, while k_{ID} indicates that rank stabilization occurs well before high-accuracy termination.

Quartic function over sphere (QS): SC holds and primal ND fails. This testbed arises from a classical polynomial optimization problem and its second-order SDP relaxation; see, *e.g.*, [59, 64]. For this relaxation, the primal solution is unique and has rank one, while primal nondegeneracy always fails. Thus these instances provide a useful testbed for the behavior of PDHG in the absence of one of the nondegeneracy conditions.

Figure 5 reports three representative instances. In all three cases, strict complementarity appears to hold numerically, as indicated by the smallest absolute eigenvalue reported above the panels. Despite the failure of primal nondegeneracy, both the KKT residual and the consecutive fixed-point difference eventually decrease R -linearly. The rank of X_k is also identified after a finite number of iterations, after which the iterates enter a clear local linear convergence regime.

Synthetic instances: SC fails and primal-dual ND holds. Figure 6 reports three synthetic SDP

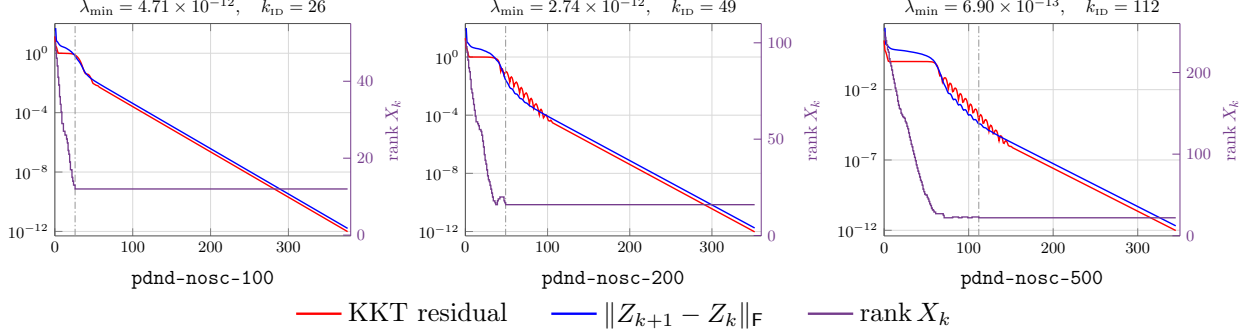


Figure 6: PDHG convergence on synthetic SDP instances. The unique KKT point satisfies primal–dual nondegeneracy and fails strict complementarity. The red and blue curves report the KKT residual and the fixed-point difference, respectively, on a logarithmic scale; the purple curve shows $\text{rank } X_k$ on the right axis in the standard scale. The dash-dotted vertical line marks the rank-identification iteration k_{ID} , and $\lambda_{\min}$ denotes the smallest absolute eigenvalue of the final Z -iterate.

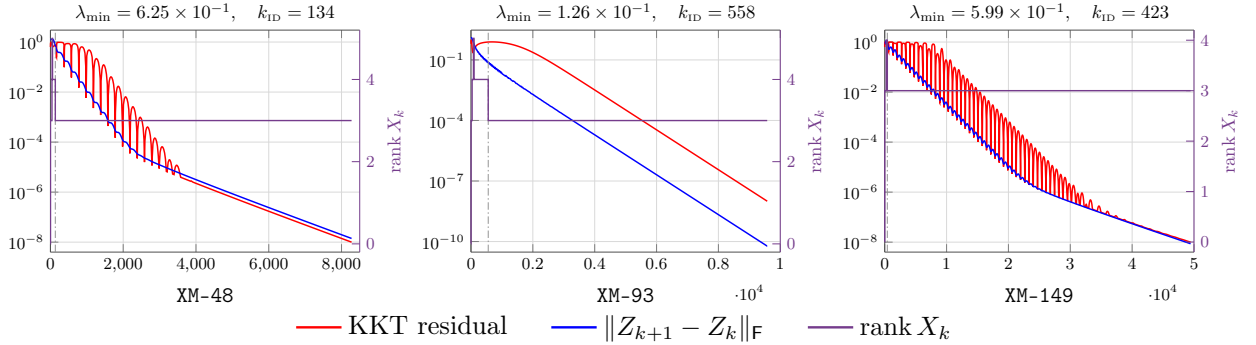


Figure 7: PDHG convergence on structure-from-motion problems. The red and blue curves report the KKT residual and the fixed-point difference, respectively, on a logarithmic scale; the purple curve shows $\text{rank } X_k$ on the right axis in the standard scale. The dash-dotted vertical line marks the rank-identification iteration k_{ID} , and $\lambda_{\min}$ denotes the smallest absolute eigenvalue of the final Z -iterate.

instances constructed so that the optimal solution is unique and satisfies primal–dual nondegeneracy, while strict complementarity fails. This is reflected in the values of $\lambda_{\min}$, which are at the level of numerical roundoff. Despite the failure of strict complementarity, PDHG exhibits a clear local linear convergence regime in all three cases. The rank of X_k is identified after a short transient phase, and thereafter both the KKT residual and the consecutive fixed-point difference decay R -linearly. These examples illustrate that strict complementarity is not necessary for the local linear behavior observed in practice, provided that the primal–dual nondegeneracy condition holds.

Structure-from-motion problems: SC holds. Figure 7 reports three SDP instances from the structure-from-motion testbed [29]. For these instances, strict complementarity appears to hold numerically. Verifying the relevant nondegeneracy conditions is computationally expensive for this testbed, so we do not report a nondegeneracy certificate. Nevertheless, the observed behavior is consistent with the preceding experiments: after a transient phase, (R) -linear convergence is observed, and the numerical rank of X_k stabilizes after finitely many iterations.

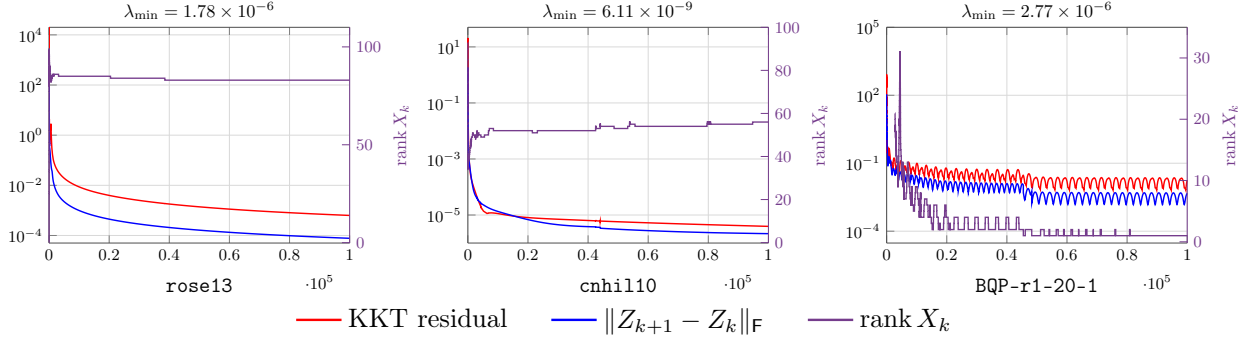


Figure 8: PDHG convergence on additional SDP instances. The red and blue curves report the KKT residual and the fixed-point difference, respectively, on a logarithmic scale; the purple curve shows $\text{rank } X_k$ on the right axis in the standard scale. Here $\lambda_{\min}$ denotes the smallest absolute eigenvalue of the final Z -iterate. For these SDPs, PDHG fails to achieve the accuracy 10^{-8} within the stated budget, and no clear linear convergence is observed.

7.2 Slow convergence near failure of strict complementarity

Figure 8 reports several SDP instances for which PDHG does not reach the prescribed accuracy within the computational budget, and for which no clear linear convergence regime is observed. A common feature of these instances is that the smallest absolute eigenvalue of the final Z -iterate is small, typically between 10^{-6} and 10^{-9} , but not numerically zero. This suggests that the corresponding instances are close to violating strict complementarity.

There are at least two possible explanations for the observed slow convergence. First, PDHG may have already entered a local linear regime, but with a contraction factor very close to one, making the progress negligible over the observed iteration horizon. Second, the local linear regime of PDHG, if present, may not yet have been reached. The latter possibility is consistent with recent observations for PDHG applied to linear programming [44]. A rigorous theoretical explanation of this behavior in SDP remains open.

8 Conclusion

We established local linear convergence of the primal–dual hybrid gradient (PDHG) algorithm for semidefinite programming (SDP) under two regularity assumptions: if the limit KKT point of PDHG satisfies strict complementarity or primal–dual nondegeneracy, then the convergence is eventually (R -)linear. The proof leverages the interpretation of PDHG as a preconditioned proximal point method. Under strict complementarity, the PSD-cone projection is smooth near the limiting signed matrix and the local KKT set has a fixed face structure. The resulting linearization contracts the component normal to the local KKT manifold. Under primal–dual nondegeneracy, a local error bound follows from strong metric subregularity of the KKT mapping. Combining these local properties with the descent property of PDHG gives the eventual linear rate. Moreover, we constructed a small SDP instance for which a formal local reduction of the PDHG map predicts sublinear convergence when both regularity conditions fail. The numerical behavior is consistent with this prediction, suggesting that the regularity assumptions used in the analysis cannot, in general, be dropped without changing the local convergence behavior.

Numerical experiments support the theory on SDP instances from different regularity regimes. They also reveal instances where PDHG has difficulty reaching high accuracy, similar to observations made recently for linear programming.

References

- [1] Jacob M. Aguirre, Diego Cifuentes, Vincent Guigues, Renato D. C. Monteiro, Victor Hugo Nascimento, and Arnesh Sujnani. cuHALLaR: A GPU accelerated low-rank augmented Lagrangian method for large-scale semidefinite programming. *arXiv preprint*, arXiv:2505.13719, 2025.
- [2] Farid Alizadeh. Interior point methods in semidefinite programming with applications to combinatorial optimization. *SIAM Journal on Optimization*, 5(1):13–51, 1995.
- [3] Farid Alizadeh, Jean-Pierre A. Haeberly, and Michael L. Overton. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming*, 77(2):111–128, 1997.
- [4] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O’Donoghue, and Warren Schudy. Practical large-scale linear programming using primal-dual hybrid gradient. In *Advances in Neural Information Processing Systems*, volume 34, pages 20243–20257, 2021.
- [5] David Applegate, Mateo Díaz, Oliver Hinder, Haihao Lu, Miles Lubin, Brendan O’Donoghue, and Warren Schudy. PDLP: A practical first-order method for large-scale linear programming. *Mathematical Programming Computation*, 2026.
- [6] David Applegate, Oliver Hinder, Haihao Lu, and Miles Lubin. Faster first-order primal-dual methods for linear programming using restarts and sharpness. *Mathematical Programming*, 201(1–2):133–184, 2023.
- [7] Sebastian Banert, Manu Upadhyaya, and Pontus Giselsson. The Chambolle–Pock method converges weakly with $\theta > 1/2$ and $\tau\sigma\|L\|^2 < 4/(1 + 2\theta)$. *Optimization Letters*, 20(3):503–520, 2026.
- [8] Heinz H. Bauschke and Patrick L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Cham, 2 edition, 2017.
- [9] Stephen P. Boyd, Laurent El Ghaoui, Eric Feron, and Venkataramanan Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*, volume 15 of *Studies in Applied Mathematics*. Society for Industrial and Applied Mathematics, Philadelphia, PA, 1994.
- [10] Stephen P. Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [11] Stephen P. Boyd and Lieven Vandenbergh. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.
- [12] Samuel Burer and Renato D. C. Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.
- [13] Samuel Burer and Renato D. C. Monteiro. Local minima and convergence in low-rank semidefinite programming. *Mathematical Programming*, 103(3):427–444, 2005.
- [14] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.
- [15] Antonin Chambolle and Thomas Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Mathematical Programming*, 159(1–2):253–287, 2016.
- [16] Zi Xian Chan and Defeng Sun. Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming. *SIAM Journal on Optimization*, 19(1):370–396, 2008.
- [17] Xiaokai Chang and Junfeng Yang. A golden ratio primal-dual algorithm for structured convex optimization. *Journal of Scientific Computing*, 87(2):47, 2021.

- [18] Kaihuang Chen, Defeng Sun, Yancheng Yuan, Guojun Zhang, and Xinyuan Zhao. HPR-LP: An implementation of an HPR method for solving linear programming. *Mathematical Programming Computation*, 2025.
- [19] Wei Deng and Wotao Yin. On the global and linear convergence of the generalized alternating direction method of multipliers. *Journal of Scientific Computing*, 66(3):889–916, 2016.
- [20] Lijun Ding, Haihao Lu, and Jinwen Yang. New understandings and computation on augmented Lagrangian methods for low-rank semidefinite programming. *arXiv preprint*, arXiv:2505.15775, 2025.
- [21] Asen L. Dontchev and R. Tyrrell Rockafellar. Regularity and conditioning of solution mappings in variational analysis. *Set-Valued Analysis*, 12(1–2):79–109, 2004.
- [22] Dmitriy Drusvyatskiy and Adrian S. Lewis. Optimality, identifiability, and sensitivity. *Mathematical Programming*, 147(1–2):467–498, 2014.
- [23] Jonathan Eckstein and Dimitri P. Bertsekas. On the Douglas–Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1–3):293–318, 1992.
- [24] Ernie Esser, Xiaoqun Zhang, and Tony 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(4):1015–1046, 2010.
- [25] Daniel Gabay and Bertrand Mercier. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers and Mathematics with Applications*, 2(1):17–40, 1976.
- [26] Pontus Giselsson and Stephen P. Boyd. Linear convergence and metric selection for Douglas–Rachford splitting and ADMM. *IEEE Transactions on Automatic Control*, 62(2):532–544, 2017.
- [27] Antoine Groudiev, Shucheng Kang, and Heng Yang. cuADMM: GPU-accelerated first-order optimization for large-scale multi-block semidefinite programs. In *2025 RSS Workshop on Fast Motion Planning and Control in the Era of Parallelism*, 2025.
- [28] Deren Han, Defeng Sun, and Liwei Zhang. Linear rate convergence of the alternating direction method of multipliers for convex composite programming. *Mathematics of Operations Research*, 43(2):622–637, 2018.
- [29] Haoyu Han and Heng Yang. Building Rome with convex optimization. In *Robotics: Science and Systems*, 2025.
- [30] Qiushi Han, Chenxi Li, Zhenwei Lin, Caihua Chen, Qi Deng, Dongdong Ge, Huikang Liu, and Yinyu Ye. A low-rank ADMM splitting approach for semidefinite programming. *INFORMS Journal on Computing*, 2025.
- [31] Bingsheng He, Feng Ma, Shengjie Xu, and Xiaoming Yuan. A generalized primal-dual algorithm with improved convergence condition for saddle point problems. *SIAM Journal on Imaging Sciences*, 15(3):1157–1183, 2022.
- [32] Bingsheng He and Xiaoming Yuan. Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective. *SIAM Journal on Imaging Sciences*, 5(1):119–149, 2012.
- [33] Xin Jiang and Lieven Vandenbergh. Bregman primal-dual first-order method and applications to sparse semidefinite programming. *Computational Optimization and Applications*, 81(1):127–159, 2022.
- [34] Xin Jiang and Lieven Vandenbergh. Bregman three-operator splitting methods. *Journal of Optimization Theory and Applications*, 196(3):936–972, 2023.

- [35] Shucheng Kang, Xin Jiang, and Heng Yang. Local linear convergence of the alternating direction method of multipliers for semidefinite programming under strict complementarity. *arXiv preprint*, arXiv:2503.20142, 2025.
- [36] Shucheng Kang and Heng Yang. Local second-order limit dynamics of ADMM for SDP. *arXiv preprint*, 2026.
- [37] Guanghui Lan and Tianjiao Li. Auto-conditioned primal-dual hybrid gradient method and alternating direction method of multipliers. *arXiv preprint*, 2024.
- [38] Jean B. Lasserre. *Moments, Positive Polynomials and Their Applications*. Imperial College Press Optimization Series. Imperial College Press, London, 2009.
- [39] Adrian S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.
- [40] Hongpei Li, Yicheng Huang, Huikang Liu, Dongdong Ge, and Yinyu Ye. D-PDLP: Scaling PDLP to distributed multi-GPU systems. *arXiv preprint*, arXiv:2601.07628, 2026.
- [41] Jingwei Liang, Jalal Fadili, and Gabriel Peyré. Local convergence properties of Douglas–Rachford and alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 172(3):874–913, 2017.
- [42] Zhenwei Lin, Zikai Xiong, Dongdong Ge, and Yinyu Ye. PDCS: A primal–dual large-scale conic programming solver with GPU enhancements. *arXiv preprint*, arXiv:2505.00311, 2025.
- [43] Haihao Lu and Jinwen Yang. cuPDLP.jl: A GPU implementation of restarted primal-dual hybrid gradient for linear programming in Julia. *Operations Research*, 73(6):3440–3452, 2025.
- [44] Haihao Lu and Jinwen Yang. On the geometry and refined rate of primal–dual hybrid gradient for linear programming. *Mathematical Programming*, 212:349–387, 2025.
- [45] Yumin Ma, Xingju Cai, Bo Jiang, and Deren Han. Understanding the convergence of the preconditioned PDHG method: A view of indefinite proximal ADMM. *Journal of Scientific Computing*, 94(3):60, 2023.
- [46] Anirudha Majumdar, Georgina Hall, and Amir Ali Ahmadi. Recent scalability improvements for semidefinite programming with applications in machine learning, control, and robotics. *Annual Review of Control, Robotics, and Autonomous Systems*, 3:331–360, 2020.
- [47] Renato D. C. Monteiro, Arnesh Sujunani, and Diego Cifuentes. A low-rank augmented Lagrangian method for large-scale semidefinite programming based on a hybrid convex-nonconvex approach. *Mathematical Programming*, 2026.
- [48] Daniel O’Connor and Lieven Vandenbergh. On the equivalence of the primal-dual hybrid gradient method and Douglas–Rachford splitting. *Mathematical Programming*, 179(1–2):85–108, 2020.
- [49] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. An algorithm for minimizing the Mumford–Shah functional. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1133–1140, 2009.
- [50] R. Tyrrell Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976.
- [51] Jos F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11(1–4):625–653, 1999.
- [52] Defeng Sun and Jie Sun. Semismooth matrix-valued functions. *Mathematics of Operations Research*, 27(1):150–169, 2002.

- [53] Tianyun Tang and Kim-Chuan Toh. A preconditioned augmented Lagrangian method for solving semidefinite programming problems. *arXiv preprint*, arXiv:2605.17089, 2026.
- [54] Michael J. Todd. Semidefinite optimization. *Acta Numerica*, 10:515–560, 2001.
- [55] Kim-Chuan Toh, Reha H. Tütüncü, and Michael J. Todd. *SDPT3 Version 3.02: A MATLAB Software for Semidefinite-Quadratic-Linear Programming*, 2002. Available at <https://www.math.cmu.edu/~reha/sdpt3.html>.
- [56] Manu Upadhyaya. The Chambolle–Pock method also converges weakly with $0 < \theta \leq 1$ and $\tau\sigma\|L\|^2 < 4\theta(2 - \theta)/(1 - 2\theta + 9\theta^2 - 4\theta^3)$. *arXiv preprint*, 2026.
- [57] Lieven Vandenbergh and Martin S. Andersen. Chordal graphs and semidefinite optimization. *Foundations and Trends in Optimization*, 1(4):241–433, 2015.
- [58] Lieven Vandenbergh and Stephen P. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.
- [59] Jie Wang and Liangbing Hu. Solving low-rank semidefinite programs via manifold optimization. *Journal of Scientific Computing*, 104(1):33, 2025.
- [60] Yinjun Wang, Haixiang Lan, and Yinyu Ye. A tuning-free primal-dual splitting algorithm for large-scale semidefinite programming. *arXiv preprint*, arXiv:2402.00311, 2024.
- [61] Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3–4):203–230, 2010.
- [62] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenbergh, editors. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*. Kluwer Academic Publishers, Boston, MA, 2000.
- [63] Stephen J. Wright. Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31(4):1063–1079, 1993.
- [64] Heng Yang, Ling Liang, Luca Carlone, and Kim-Chuan Toh. An inexact projected gradient method with rounding and lifting by nonlinear programming for solving rank-one semidefinite relaxation of polynomial optimization. *Mathematical Programming*, 201(1–2):409–472, 2023.
- [65] Liuqin Yang, Defeng Sun, and Kim-Chuan Toh. SDPNAL+: A majorized semismooth Newton-CG augmented Lagrangian method for semidefinite programming with nonnegative constraints. *Mathematical Programming Computation*, 7(3):331–366, 2015.
- [66] Xinyuan Zhao, Defeng Sun, and Kim-Chuan Toh. A Newton-CG augmented Lagrangian method for semidefinite programming. *SIAM Journal on Optimization*, 20(4):1737–1765, 2010.