Local Linear Convergence of the Alternating Direction Method of Multipliers for Semidefinite Programming under Strict Complementarity

Shucheng Kang[*] Xin Jiang[†] Heng Yang[‡]

April 14, 2025

Abstract

We investigate the local linear convergence properties of the Alternating Direction Method of Multipliers (ADMM) when applied to Semidefinite Programming (SDP). A longstanding belief suggests that ADMM is only capable of solving SDPs to moderate accuracy, primarily due to its sublinear worst-case complexity and empirical observations of slow convergence. We challenge this notion by introducing a new sufficient condition for local linear convergence: as long as the converged primal–dual optimal solutions satisfy strict complementarity, ADMM attains local linear convergence, independent of nondegeneracy conditions. Our proof is based on a direct local linearization of the ADMM operator and a refined error bound for the projection onto the positive semidefinite cone, improving previous bounds and revealing the anisotropic nature of projection residuals. Extensive numerical experiments confirm the significance of our theoretical results, demonstrating that ADMM achieves local linear convergence and computes high-accuracy solutions in a variety of SDP instances, including those where nondegeneracy fails. Furthermore, we identify cases where ADMM struggles, linking these difficulties with near violations of strict complementarity—a phenomenon that parallels recent findings in linear programming.

[*]School of Engineering and Applied Sciences, Harvard University. Email: skang1@g.harvard.edu
[†]School of Operations Research and Information Engineering, Cornell University. Email: xjiang@cornell.edu
[‡]School of Engineering and Applied Sciences, Harvard University. Email: hankyang@seas.harvard.edu

Contents

1 Introduction

Consider the semidefinite programs (SDPs) in the standard form:

$$\begin{array}{llllll} \text{Primal:} & \text{minimize} & \langle C, X\rangle & \quad \text{Dual:} & \text{maximize} & b^{\mathsf{T}} y \\ & \text{subject to} & \mathcal{A}X = b & & \text{subject to} & \mathcal{A}^* y + S = C \\ & & X \in \mathbb{S}^n_+ & & & S \in \mathbb{S}^n_+, \end{array} \tag{1}$$

with primal variable $X \in \mathbb{S}^n$ and dual variables $S \in \mathbb{S}^n$, $y \in \mathbb{R}^m$, where $\mathbb{S}^n$ is the set of real symmetric $n \times n$ matrices and $\mathbb{S}^n_+$ is the set of positive semidefinite (PSD) matrices in $\mathbb{S}^n$. The linear operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is defined as

$$\mathcal{A}X := (\langle A_1, X\rangle, \cdots, \langle A_m, X\rangle)$$

and $\mathcal{A}^* y = \sum_{i=1}^m y_i A_i$ is its adjoint operator. The coefficients $C, A_1, \ldots, A_m$ are symmetric $n \times n$ matrices. It is assumed that $\{A_i\}_{i=1}^m$ are linearly independent so that $\mathcal{A}\mathcal{A}^*$ is an invertible operator.

With the growing demand for solving large-scale SDPs, particularly those arising from moment and sums-of-squares (SOS) relaxations in polynomial optimization [19,21,28,30–32,41,43,57,62], first-order methods (FOMs) have gained increasing attention due to their low per-iteration cost and ability to exploit problem structure. Among these, the Alternating Direction Method of Multipliers (ADMM) has emerged as a widely adopted approach, with numerous implementations, applications, and variations [11,31,46,58,64,69].

ADMM for SDP. Starting from $(X^{(0)}, y^{(0)}, S^{(0)})$, the classical three-step ADMM iteration for the SDP (1) reads as [58]:

$$y^{(k+1)} = (\mathcal{A}\mathcal{A}^*)^{-1} \left(\sigma^{-1} b - \mathcal{A}\left(\sigma^{-1} X^{(k)} + S^{(k)} - C\right)\right) \tag{2a}$$

$$S^{(k+1)} = \Pi_{\mathbb{S}^n_+} \left(C - \mathcal{A}^* y^{(k+1)} - \sigma^{-1} X^{(k)}\right) \tag{2b}$$

$$X^{(k+1)} = X^{(k)} + \sigma \left(S^{(k+1)} + \mathcal{A}^* y^{(k+1)} - C\right) \tag{2c}$$

where $\Pi_{\mathbb{S}^n_+}(\cdot)$ denotes the orthogonal projection onto the PSD cone $\mathbb{S}^n_+$ and $\sigma > 0$ is the penalty parameter. Under mild conditions, $(X^{(k)}, y^{(k)}, S^{(k)})$ is convergent to $(X_\star, y_\star, S_\star)$, one of the optimal solution pairs satisfying the Karush–Kuhn–Tucker (KKT) conditions [58, Theorem 2]:

$$\mathcal{A}X_\star = b, \qquad \mathcal{A}^* y_\star + S_\star = C, \qquad \langle X_\star, S_\star\rangle = 0, \qquad X_\star \in \mathbb{S}^n_+, \qquad S_\star \in \mathbb{S}^n_+. \tag{3}$$

The ADMM iteration (2) is often analyzed via the following equivalent fixed-point iterations [34]

$$\begin{aligned} Z^{(k+1)} = {} & \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(-2\Pi_{\mathbb{S}^n_+}(Z^{(k)}) + Z^{(k)}) + \Pi_{\mathbb{S}^n_+}(Z^{(k)}) \\ & + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1} b + \sigma \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}C - \sigma C, \end{aligned} \tag{4}$$

where we make the change of variables $Z := X - \sigma S$ (and $Z_\star := X_\star - \sigma S_\star$). From $Z^{(k)}$, we can extract the primal variable and the (scaled) dual variable as

$$X^{(k)} = \Pi_{\mathbb{S}^n_+}(Z^{(k)}), \qquad \sigma S^{(k)} = \Pi_{\mathbb{S}^n_+}(-Z^{(k)}). \tag{5}$$

In this paper, we investigate the local convergence properties of ADMM for solving SDPs. To this end, we begin by recalling two important regularity conditions in SDP.

Nondegeneracy and strict complementarity. First introduced in [1], nondegeneracy and strict complementarity have been two fundamental regularity conditions in SDP [2,68]. Since the primal–dual optimal solutions are simultaneously diagonalizable [59, pp. 308], we assume they admit the following decomposition:

$$X_\star = Q_\star \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix} Q_\star^\mathsf{T}, \quad \Lambda_X := \mathrm{diag}\left(\lambda_1, \ldots, \lambda_r\right) \tag{6a}$$

$$\sigma S_\star = Q_\star \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_S \end{bmatrix} Q_\star^\mathsf{T}, \quad \Lambda_S := -\mathrm{diag}\left(\lambda_{n-s+1}, \ldots, \lambda_n\right), \tag{6b}$$

where $\mathrm{diag}\left(\cdot\right)$ assembles a vector into a diagonal matrix, $Q_\star \in \mathbb{R}^{n \times n}$ is an orthogonal matrix, and the eigenvalues satisfy

$$\lambda_1 \geq \cdots \geq \lambda_r > 0 > \lambda_{n-s+1} \geq \cdots \geq \lambda_n$$

with $r + s \leq n$. Then, we define four important subspaces [1]:

$$\mathcal{T}_{X_\star} := \left\{ Q_\star \begin{bmatrix} H_X & H_O^\mathsf{T} \\ H_O & 0 \end{bmatrix} Q_\star^\mathsf{T} \;\middle|\; H_X \in \mathbb{S}^r,\, H_O \in \mathbb{R}^{(n-r)\times r} \right\} \tag{7a}$$

$$\mathcal{N}_{X_\star} := \mathcal{T}_{X_\star}^\perp = \left\{ Q_\star \begin{bmatrix} 0 & 0 \\ 0 & H_S \end{bmatrix} Q_\star^\mathsf{T} \;\middle|\; H_S \in \mathbb{S}^{n-r} \right\} \tag{7b}$$

$$\mathcal{T}_{S_\star} := \left\{ Q_\star \begin{bmatrix} 0 & H_O^\mathsf{T} \\ H_O & H_S \end{bmatrix} Q_\star^\mathsf{T} \;\middle|\; H_S \in \mathbb{S}^s,\, H_O \in \mathbb{R}^{s\times(n-s)} \right\} \tag{7c}$$

$$\mathcal{N}_{S_\star} := \mathcal{T}_{S_\star}^\perp = \left\{ Q_\star \begin{bmatrix} H_X & 0 \\ 0 & 0 \end{bmatrix} Q_\star^\mathsf{T} \;\middle|\; H_X \in \mathbb{S}^{n-s} \right\}, \tag{7d}$$

where $\mathcal{L}^\perp$ represents the orthogonal complement of the linear subspace $\mathcal{L}$. Further, we denote $\mathcal{R}(\mathcal{A}^*) := \{\sum_{i=1}^m A_i y_i \mid y \in \mathbb{R}^m\}$ as the range space of $\mathcal{A}^*$, and $\mathcal{N}(\mathcal{A})$ as $\mathcal{R}(\mathcal{A}^*)^\perp = \{X \in \mathbb{S}^n \mid \mathcal{A}X = 0\}$. The nondegeneracy (ND) and strict complementarity (SC) conditions—two generic properties for SDPs—are defined as [1]:

$$\text{Primal Nondegeneracy: } \mathcal{T}_{X_\star} + \mathcal{N}(\mathcal{A}) = \mathbb{S}^n \iff \mathcal{N}_{X_\star} \cap \mathcal{R}(\mathcal{A}^*) = \{0\} \tag{8}$$

$$\text{Dual Nondegeneracy: } \mathcal{T}_{S_\star} + \mathcal{R}(\mathcal{A}^*) = \mathbb{S}^n \iff \mathcal{N}_{S_\star} \cap \mathcal{N}(\mathcal{A}) = \{0\} \tag{9}$$

$$\text{Strict Complementarity: } \mathrm{rank}(X_\star) + \mathrm{rank}(S_\star) = n \iff r + s = n. \tag{10}$$

When strict complementarity holds, primal (resp., dual) nondegeneracy is equivalent to the uniqueness of dual (resp., primal) optimal solution [1].

1.1 ADMM for SDP: Sublinear Rate and Moderate Accuracy?

A common perception for ADMM among practitioners is that it generally cannot solve the SDP (1) to high accuracy (*e.g.,* max KKT residual below 10^{-10}). Indeed, this perception arises from both theoretical challenges and empirical observations.

Theoretically, a well-accepted convergence rate for ADMM applied to solving SDPs is $\mathcal{O}\left(1/\epsilon\right)$, which matches the general convex optimization case [51]. Establishing linear convergence $\mathcal{O}(\log(1/\epsilon))$ is considered hard, since known sufficient conditions, such as strong convexity [44], local polyhedrality [27, 35] and certain growth conditions [13, 15, 27, 39, 65, 66], either fail or remain unclear for SDP. As pioneering works, [9, 22] established local linear convergence of ADMM for solving SDPs when primal nondegeneracy (8) and dual nondegeneracy (9) both hold (regardless of strict complementarity (10)), by showing the metric subregularity (or calmness) of the KKT operator. However, two challenges remain: (a) two-side nondegeneracy conditions are hard to check numerically, and (b) important subclasses of SDP, such as those arising from the moment-SOS relaxations with finite convergence [32], are known to be degenerate. It remains an open question in optimization theory whether alternative, and perhaps simple-to-verify, sufficient conditions can ensure the linear convergence of ADMM for solving SDPs.

Empirically, the slow convergence of ADMM and its variants when solving SDPs is widely reported [20, 31, 63, 69]. For a typical SDP, the interior point method [2] (IPM) implemented in MOSEK [4] can solve the problem to machine precision (if memory permits), while ADMM often struggles to achieve moderate accuracy (e.g., max KKT residual 10^{-4}). Consequently, ADMM and its variants are primarily used to warmstart downstream solvers [63, 64] or as efficient methods to obtain coarse solutions [31]. In contrast, recently in the linear programming (LP) literature, first-order methods are observed to exhibit significantly improved local linear convergence rates after initially traversing a prolonged phase of slow convergence [40]. To the best of our knowledge, no comparable numerical evidence has been documented in the SDP literature.

Thus, the central question driving this paper is:

> Can ADMM exhibit empirically observable linear convergence when solving SDPs? If so, can we establish numerically verifiable sufficient conditions to guarantee this behavior?

1.2 Contributions

We affirm this question both empirically and theoretically. We establish local linear convergence of ADMM for SDP under a mild strict complementarity (SC) assumption. Moreover, comprehensive numerical results are reported to demonstrate such a prevalent linear rate of convergence.

Theoretical contribution. We establish a new sufficient condition that guarantees the local linear convergence of ADMM for solving SDPs.

Theorem 1 (Informal: Local Linear Convergence under Strict Complementarity)**.** *If ADMM converges to an optimal solution $(X_\star, y_\star, S_\star)$ of the SDP* (1) *that satisfies strict complementarity* (10)*, then ADMM attains local linear convergence after finite iterations.*

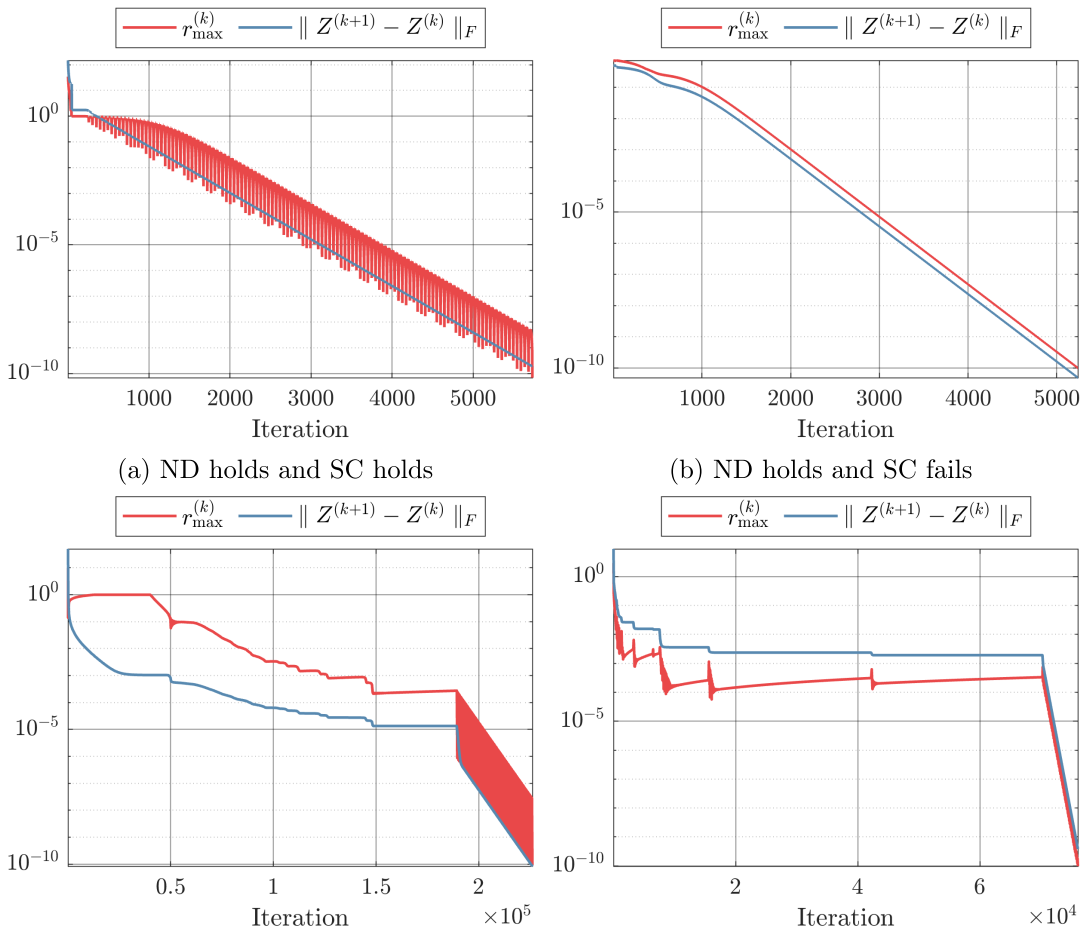
Our sufficient condition holds independently of nondegeneracy (ND). Verifying strict complementarity (10) requires only checking the numerical rank of $X_\star$ and $S_\star$, a significantly more tractable procedure compared with verifying nondegeneracy, which involves examining the intersection of two subspaces.

At a high level, our proof framework is built upon a detailed local linearization analysis of the ADMM operator, incorporating several key contributions.

- **A refined error bound for the PSD cone projector.** We improve the classic linearization result of the PSD cone projector in [54, Theorem 4.6] by tightening the bound on the linearization residual from $\mathcal{O}(\|H\|^2)$ to $\mathcal{O}(\|H_O\| \cdot \|H\|)$, where H represents a small perturbation and H_O denotes its off-block-diagonal part. This refinement is a cornerstone of our proof, revealing the anisotropic nature of the residual and offering potential applications beyond our setting.
- **Characterization of the local behavior of ADMM.** Our analysis relies on a local linearization of the fixed-point iteration (9). In particular, when near optimum, the iterate error $Z^{(k+1)} - Z_\star$ can be written as a linear transformation of $Z^{(k)} - Z_\star$ plus a quadratic residual term. When both ND and SC hold, we prove that the linear transformation is contractive and that the linearization residual becomes negligible, directly leading to local linear convergence.
- **Handling cases without nondegeneracy.** When ND fails but SC holds, we first show that the several "partial" sequences, *e.g.,* the off-block-diagonal part of $Z^{(k)} - Z_\star$, vanish at a local linear rate. To establish the convergence of the full sequence, we further extend the regularized backward error bound for spectrahedra [53, Lemma 2.3] to the (scaled) KKT system of SDP. This helps build a local "conditional sharpness" property, analogous to sharpness in LP [3] but with an additional term related to the minimal faces of the PSD cone. This "conditional sharpness", together with the convergence of "partial" sequences, finally yields the R-linear rate of convergence for the ADMM iterates.

Notably, our proof framework differs from [22] in that it does not rely on the metric subregularity of the KKT operator, which is intractable to verify numerically.

Empirical contribution. In parallel with the theoretical findings, we present extensive numerical evidence demonstrating the prevalent local linear convergence of ADMM when solving SDPs. To systematically analyze this phenomenon, we categorize SDP problems into four families based on whether the nondegeneracy (ND) and strict complementarity (SC) conditions hold or fail. A representative subset of our numerical results is shown in Figure 1, where ADMM consistently enters a linear convergence regime across all four cases. The full set of numerical experiments is detailed in Section 8, covering a broad range of SDP instances. Our test suite includes standard benchmark datasets [42, 48] as well as newly generated SDP problems in real-world applications [23, 62]. These instances span classical MAXCUT-style SDPs and more challenging problems from the moment-SOS relaxations with finite convergence [32, 56]. All the SDP problems are available at https://github.com/ComputationalRobotics/admmsdp-linearconv.



(a) ND holds and SC holds

(b) ND holds and SC fails

(c) ND fails and SC holds

(d) ND fails and SC fails

Figure 1: Four representative SDP instances. (a) A toy structure-from-motion problem from [23]; (b) A toy example from [59, pp. 44]; (c) A Quasar problem from [61] with random initialization; (d) Second-order relaxation for a random BQP problem [63] with all-zeros initialization. Here, $r_{\max}$ denotes the maximum KKT residual. In all the cases, ADMM with fixed σ parameter eventually exhibits local linear convergence.

In addition, we document "failure" cases of ADMM in which the maximum KKT residual remains above 10^{-10} even after reaching the iteration limit (10^6 iterations) or the time limit (100 hours). These cases exhibit a common feature: the minimum positive eigenvalue of the converged $X_\star$ or $S_\star$ is near zero. This behavior closely resembles difficult cases in LP [40] and can be partly explained via our proof framework.

Open questions: rank identification and beyond. In first-order methods for LP [40], local linear convergence is achieved alongside with basis identification. An analogous result in ADMM for SDP (with SC) would be *rank identification*; *i.e.,* after a finite number of iterations, ADMM identifies the solution rank, and all subsequent iterates maintain at the same rank. Though this result could be readily drawn from the partial smoothness theory [18, 33, 60], we provide a more direct proof that only leverages the algorithmic properties of ADMM. However, unlike the LP case, it remains unclear that whether rank identification and linear convergence occur *simultaneously* in ADMM for SDP. In this work, we provide a partial answer through a numerical example and leave a full investigation for future work.

Relating our discussion back to Figure 1, this work establishes (R-)linear convergence guarantees of ADMM for SDP, which covers cases (a)–(c) in Figure 1. (In comparison, [22] explains cases (a) and (b).) However, whether ADMM provably attains local linear convergence in case (d), where both ND and SC fail, remains an open question, a gap between theory and practice that warrants further investigation.

1.3 Notations

We use $\mathbb{R}^n$ to denote the set of n-dimensional real vectors and $\mathbb{R}^n_+$ (resp., $\mathbb{R}^n_{++}$) the set of nonnegative (resp., positive) vectors in $\mathbb{R}^n$. Denote $\mathbb{R}^{m\times n}$ as the set of $m \times n$ matrices. Denote $\mathbb{S}^n$ as the set of real symmetric $n \times n$ matrices and $\mathbb{S}^n_+/\mathbb{S}^n_{++}$ (resp. $\mathbb{S}^n_-/\mathbb{S}^n_{--}$) as the set of positive semidefinite/positive definite matrices (resp., negative semidefinite/negative definite matrices) in $\mathbb{S}^n$. Denote $\mathbb{N}$ the set of nonnegative integers and for any integer $n \in \mathbb{N}$, define $[n] := 1, 2, \dots, n$. Denote Id as the identity operator, denote I_n as the $n \times n$ identity matrix and E_n (resp., $E_{m\times n}$) as the $n \times n$ (resp. $m \times n$) all-ones matrix. For $A \in \mathbb{S}^n$, $\lambda_{\min}(A)$ (resp., $\lambda_{\max}(A)$) represents its minimal eigenvalue (resp., maximal eigenvalue). For $x \in \mathbb{R}^n$, we denote $\|x\|_2$ as its Euclidean norm. For $X \in \mathbb{R}^{m\times n}$, $\|X\|_2$ represents its spectral norm, $\|X\|_{\mathsf{F}}$ its Frobenius norm, and $\|X\|$ an arbitrary norm. For a linear operator $\mathcal{M} : \mathbb{S}^n \to \mathbb{S}^n$, we use $\|\mathcal{M}\|_{\text{op}}$ to denote its operator norm: $\|\mathcal{M}\|_{\text{op}} := \sup\{\|\mathcal{M}X\|_{\mathsf{F}} \mid \|X\|_{\mathsf{F}} = 1\}$.

Denote $A \circ B$ as the Hadamard product between two matrices A and B of the same size; *i.e.,* $(A \circ B)_{ij} = A_{ij}B_{ij}$. Denote $A \otimes B$ as the Kronecker product between $A \in \mathbb{R}^{m\times n}$ and $B \in \mathbb{R}^{p\times q}$, and denote $A \oplus B = A \otimes I_m + I_n \otimes B$ as the Kronecker sum between $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{m\times m}$. Denote $\operatorname{vec}(X)$ has the column-major vectorization of an arbitrary matrix X, and denote $\operatorname{svec}(A) : \mathbb{S}^n \to \mathbb{R}^{t(n)}$ as the symmetric vectorization of A, where $t(n) := \frac{n(n+1)}{2}$. Denote smat as the inverse operator of svec. For an arbitrary matrix A, $A_{a:b,c:d}$ represents the submatrix of A indexed from row a to row b and from column c to column d.

The distance from a point $X \in \mathbb{S}^n$ to a set $\mathcal{X} \subseteq \mathbb{S}^n$ is defined as

$$\operatorname{dist}(X, \mathcal{X}) := \inf_{\widetilde{X}\in\mathcal{X}} \|X - \widetilde{X}\|_{\mathsf{F}}.$$

We will use the same notation to denote the distance to a Cartesian product of sets:

$$\operatorname{dist}((X, S), \mathcal{X} \times \mathcal{S}) := \inf_{(\widetilde{X},\widetilde{S})\in\mathcal{X}\times\mathcal{S}} \sqrt{\|X - \widetilde{X}\|_{\mathsf{F}}^2 + \|S - \widetilde{S}\|_{\mathsf{F}}^2}.$$

We denote the orthogonal projection onto a set $\mathcal{X}$ as $\Pi_{\mathcal{X}}$; in particular, $\Pi_{\mathbb{S}^n_+}(\cdot)$ means the orthogonal projection onto the PSD cone.

1.4 Outline

After a brief review of related work in Section 2, we introduce our refined error bound in Section 3, a fundamental result that underpins our proof framework and holds independent interest. We then examine the local linearization of ADMM in Section 4 and establish its local linear convergence both with and without nondegeneracy in Section 5 and Section 6, respectively. Due to its mathematical complexity, the full proof of our refined error bound is deferred to Section 7. In Section 8, we conduct extensive numerical experiments to support our theoretical findings. Section 9 briefly discusses the rank identification phenomenon as well as its relationship with local linear convergence. Finally, Section 10 includes concluding remarks.

2 Related Work

With the rapid development of data science and all fields in engineering, SDP problems are growing in scale. Efficient and scalable algorithms have been developed, analyzed and implemented for solving large-scale SDPs.

Augmented Lagrangian method (ALM). Originally introduced to enhance the performance of penalty methods [49], ALM has shown promise in tackling large-scale SDPs [64]. Under mild conditions (strong duality and the existence of a strictly complementary solution pair), linear convergence of ALM [12, 37, 68] is established by leveraging its connection to proximal point methods [50] (PPM) and quadratic growth properties [17, 53].

Burer–Monteiro (BM) factorization method. The BM method [8] replaces the conic constraint by $X = RR^\mathsf{T}$, reducing the problem to a lower-dimensional nonlinear program. Under specific rank and regularity conditions, it recovers the global optimum of the original SDP [7, 55]. Owing to its efficiency, the BM method has achieved significant empirical success in real-world problems with low-rank solutions [23]. It can also be combined with ALM [56] or ADMM [24].

Spectral bundle method (SBM). First proposed in [25], SBM has gained attention for its low per-iteration cost. It enjoys sublinear convergence under mild assumptions [16]. Furthermore, if a strictly complementary solution pair exists and the surrogate function captures the correct rank of the optimal solution, SBM achieves local linear convergence [16, 36]. Similar to ALM, its linear convergence guarantees rely on quadratic growth. Recent work also incorporates SBM into ALM [38].

First-order proximal methods. As a broad class of first-order methods derived from the monotone operator theory [51], primal–dual proximal methods are also popular for SDP. In addition to ADMM [58], symmetric Gauss-Seidel (sGS)-ADMM [11] has gained traction for solving general SDPs to medium accuracy. Its connection to proximal ALM is elaborated in [10]. Other proximal methods, such as the primal–dual hybrid gradient (PDHG) method [29], have likewise been explored. Nonetheless, establishing local linear convergence remains much harder for this class of algorithms than for ALM or SBM, because it is not yet clear whether a suitable growth condition holds for generic SDPs [22]. In particular, the known sufficient conditions for linear convergence of ADMM, *e.g.,* strong convexity [44], local polyhedrality [35] and other growth conditions [13, 15, 27, 39, 65, 66], either fail or remain unclear for SDP.

3 A Refined Error Bound for PSD Cone Projection

We see from (4) that the only nonlinear operation in ADMM is the projection onto the PSD cone. So, to better understand the convergence of ADMM for SDP, we need to study the local behavior of the PSD cone projector $\Pi_{\mathbb{S}^n_+}$. A classic result on the perturbation theory of $\Pi_{\mathbb{S}^n_+}$ is [54, Theorem 4.6], which is restated below for self-containment.

Lemma 1 ([54, Theorem 4.6]). *Given an $n \times n$ symmetric nonsingular matrix $Z \in \mathbb{S}^n$, denote its eigenvalue decomposition by*

$$Z = Q\mathrm{diag}\left(\lambda_1, \ldots, \lambda_r, \lambda_{r+1}, \ldots, \lambda_n\right) Q^\mathsf{T}, \quad \textit{where } \lambda_1 \geq \cdots \geq \lambda_r > 0 > \lambda_{r+1} \geq \cdots \geq \lambda_n$$

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

$$(\Pi_{\mathbb{S}^n_+}(Z))'(H) = Q(\Omega \circ (Q^\mathsf{T} H Q))Q^\mathsf{T},$$

where the $n \times n$ *symmetric matrix* Ω *is defined as*

$$\Omega = \begin{bmatrix} 1 & \cdots & 1 & \frac{\lambda_1}{\lambda_1-\lambda_{r+1}} & \cdots & \frac{\lambda_1}{\lambda_1-\lambda_n} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cdots & 1 & \frac{\lambda_r}{\lambda_r-\lambda_{r+1}} & \cdots & \frac{\lambda_r}{\lambda_r-\lambda_n} \\ \frac{\lambda_1}{\lambda_1-\lambda_{r+1}} & \cdots & \frac{\lambda_r}{\lambda_r-\lambda_{r+1}} & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\lambda_1}{\lambda_1-\lambda_n} & \cdots & \frac{\lambda_r}{\lambda_r-\lambda_n} & 0 & \cdots & 0 \end{bmatrix} := \begin{bmatrix} E_r & \Theta^{\mathsf{T}} \\ \Theta & 0 \end{bmatrix}. \tag{11}$$

Here, E_r *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* Ω*:*

$$\Theta_{ij} = \frac{\lambda_j}{\lambda_j - \lambda_{i+r}} \in (0,1), \quad \textit{for } i \in [n-r],\ j \in [r]. \tag{12}$$

Moreover, for any sufficiently small perturbation $H \in \mathbb{S}^n$*, it holds that*

$$\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - Q(\Omega \circ (Q^{\mathsf{T}} H Q))Q^{\mathsf{T}}\|_2 = \mathcal{O}(\|H\|_2^2).$$

However, the following simple example illustrates that the above result may not be tight and motivates our refined error bound. To see this, assume for brevity that $Q = I$ in Lemma 1 and partition the perturbation $H \in \mathbb{S}^n$ as

$$H = \begin{bmatrix} H_X & H_O^{\mathsf{T}} \\ H_O & H_S \end{bmatrix}, \quad \text{where } H_X \in \mathbb{S}^r, H_S \in \mathbb{S}^{n-r}, \text{ and } H_O \in \mathbb{R}^{(n-r)\times r}. \tag{13}$$

We set $H_O = 0$ and $\|H\|_2 \le \sigma_{\min}(Z) := \min\{\lambda_r, -\lambda_{r+1}\}$. Then we have

$$Z + H = \begin{bmatrix} \Lambda_X + H_X & 0 \\ 0 & \Lambda_S + H_S \end{bmatrix},$$

where $\Lambda_X = \operatorname{diag}(\lambda_1, \ldots, \lambda_r)$ and $\Lambda_S = \operatorname{diag}(\lambda_{r+1}, \ldots, \lambda_n)$. We then obtain from Weyl's inequality that

$$\begin{aligned} \lambda_{\min}(\Lambda_X + H_X) &\ge \lambda_r + \lambda_{\min}(H_X) \ge \lambda_r - \|H_X\|_2 \ge 0, \\ \lambda_{\max}(\Lambda_S + H_S) &\le \lambda_{r+1} + \lambda_{\max}(H_S) \le \lambda_{r+1} + \|H_S\|_2 \le 0, \end{aligned}$$

where we also use the facts that $\|H_X\|_2 \le \|H\|_2$ and $\|H_S\|_2 \le \|H\|_2$. Therefore,

$$\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) = \begin{bmatrix} \Lambda_X + H_X & 0 \\ 0 & 0 \end{bmatrix} - \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix} = \Omega \circ H,$$

i.e., the residual term is exactly zero while $\|H\|_2$ is nonzero. This motivates the following refined error bound for the PSD projection $\Pi_{\mathbb{S}^n_+}$, which involves the "off-block-diagonal" part H_O in the residual.

Theorem 2. *Given an* $n \times n$ *symmetric nonsingular matrix* $Z \in \mathbb{S}^n$*, denote its eigenvalue decomposition by*

$$Z = Q\operatorname{diag}(\lambda_1, \ldots, \lambda_r, \lambda_{r+1}, \ldots, \lambda_n)\, Q^{\mathsf{T}}, \quad \textit{where } \lambda_1 \ge \cdots \ge \lambda_r > 0 > \lambda_{r+1} \ge \cdots \ge \lambda_n$$

and $Q \in \mathbb{R}^{n\times n}$ *is an orthogonal matrix. Then, there exist two positive constants* C_{EB} *and* α_{EB} *such that for all* $H \in \mathbb{S}^n$ *with* $\|H\|_2 \le C_{\mathrm{EB}}$*, it holds that*

$$\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - Q(\Omega \circ \widetilde{H})Q^{\mathsf{T}}\|_2 \le \alpha_{\mathrm{EB}} \cdot \|\widetilde{H}_O\|_2 \cdot \|H\|_2, \tag{14}$$

where $\widetilde{H} := Q^{\mathsf{T}} H Q$ *is partitioned as*

$$\widetilde{H} = \begin{bmatrix} \widetilde{H}_X & \widetilde{H}_O^{\mathsf{T}} \\ \widetilde{H}_O & \widetilde{H}_S \end{bmatrix} \quad \textit{with } \widetilde{H}_X \in \mathbb{S}^r,\ \widetilde{H}_S \in \mathbb{S}^{n-r},\ \textit{and } \widetilde{H}_O \in \mathbb{R}^{(n-r)\times r}.$$

Remark 1. *When $Q = I$, the bound (14) reduces to*

$$\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H\|_2 \le \alpha_{\mathrm{EB}} \cdot \|H_O\|_2 \cdot \|H\|_2.$$

This aligns with our observation in the motivating example: when $H_O = 0$, both sides of the above inequality becomes zero. One shall also note that, in general, $\widetilde{H}_O$ is not the off-block-diagonal part of the perturbation H. In fact, without using the notation $\widetilde{H}$, the bound (14) can be written as

$$\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - Q(\Omega \circ (Q^\mathsf{T} H Q))Q^\mathsf{T}\|_2 \le \alpha_{\mathrm{EB}} \cdot \|Q_S^\mathsf{T} H Q_X\|_2 \cdot \|H\|_2,$$

where we partition the eigenvalue decomposition of Z as

$$Z = \begin{bmatrix} Q_X & Q_S \end{bmatrix} \begin{bmatrix} \Lambda_X & 0 \\ 0 & \Lambda_S \end{bmatrix} \begin{bmatrix} Q_X^\mathsf{T} \\ Q_S^\mathsf{T} \end{bmatrix}.$$

Remark 2. *As all norms are equivalent, (14) implies that there exists a positive constant α'_{EB} such that*

$$\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - Q(\Omega \circ \widetilde{H})Q^\mathsf{T}\|_\mathsf{F} \le \alpha'_{\mathrm{EB}} \cdot \|\widetilde{H}_O\|_\mathsf{F} \cdot \|H\|_\mathsf{F}.$$

In the convergence analysis (Sections 4 to 6), we mainly use the Frobenius norm of matrices, consistent with most literature.

Remark 3. *In [12, Proposition 3.4], the authors establish another perturbation property for the PSD cone projector. Their result has two key distinctions from Theorem 2: (1) their results cover cases where Z is singular, while ours only focuses on the nonsingular case; (2) under the nonsingularity assumption, our results can directly lead to theirs;* i.e., *Theorem 2 is stronger than [12, Proposition 3.4]. See Appendix A for detailed discussion.*

4 Local Linearization of ADMM

With a better understanding of the PSD cone projection, we now study the local behavior of ADMM. Our analysis is different from the standard approaches for ADMM and starts by locally linearizing the iteration (4). In particular, we show that when near optimum, the residual $H^{(k+1)} := Z^{(k)} - Z_\star$ is almost a linear transformation of of the previous residual $H^{(k)}$, plus a correction term in the order of $\mathcal{O}(\|H_O\|_\mathsf{F}\|H\|_\mathsf{F})$.

The rest of the section is devoted to study this linearization and is organized as follows. Section 4.1 lists all the assumptions made throughout the paper, Section 4.2 presents the local linearization of ADMM, and Section 4.3 describes the properties of such a linearization.

4.1 Assumptions

We make the following assumption on the pair of primal–dual SDPs (1).

Assumption 1. *(a) The linear operator $\mathcal{A} : \mathbb{S}^n \to \mathbb{R}^m$ is surjective.*

(b) The pair of primal–dual SDPs (1) has a nonempty set of KKT points.

From convex duality theory, any pair of primal–dual solutions $(X_\star, y_\star, S_\star)$ of (1) satisfies complementary slackness; *i.e.*,, $X_\star$ and $S_\star$ admit the decompositions in (6). Moreover, $\langle X_\star, S_\star\rangle = 0$ and $\mathrm{rank}(X_\star) + \mathrm{rank}(S_\star) \le n$. When the above inequality holds with equality, the solution pair $(X_\star, y_\star, S_\star)$ is called strictly complementary.

It is known that under Assumption 1, three-step ADMM (2) converges to a KKT point $(X_\star, y_\star, S_\star)$, or equivalently, one-step ADMM (4) converges to the point $Z_\star := X_\star - \sigma Z_\star$. Our analysis assumes additionally that the convergent point of ADMM is strictly complementary.

Assumption 2. *Three-step ADMM* (2) *converges to a KKT point* $(X_\star, y_\star, S_\star)$ *satisfying strict complementarity;* i.e.,, $\operatorname{rank}(X_\star) + \operatorname{rank}(S_\star) = n$.

Assumption 2 is equivalent to the condition that the convergent point $Z_\star$ of one-step ADMM (4) is nonsingular. Assumption 2 is a mild assumption in the sense that it holds for generic SDPs [1]. Numerical experiments in Section 8 further demonstrate that even for degenerate SDPs with multiple solutions, one-step ADMM often converges to a nonsingular $Z_\star$ (if one exists) when initialized with a random (standard Gaussian) guess $Z^{(0)}$.

For ease of presentation, we assume without loss of generality that the convergent points $X_\star$ and $S_\star$ are diagonal, *i.e.,* $Q_\star = I_n$ in (6). This assumption does not limit the scope of our conclusions because we can readily construct a pair of SDPs equivalent to (1) and generate ADMM iterates $(\widetilde{X}^{(k)}, \widetilde{y}^{(k)}, \widetilde{S}^{(k)})$ orthogonally similar to the iterates $(X^{(k)}, y^{(k)}, S^{(k)})$ generated by (2). To see this, suppose $(X^{(k)}, y^{(k)}, S^{(k)})$ converges to $(X_\star, y_\star, S_\star)$ with $Q_\star \neq I_n$. We construct another pair of SDPs

$$\begin{array}{llll} \text{Primal:} & \text{minimize} & \langle \widetilde{C}, \widetilde{X}\rangle & \qquad \text{Dual:} & \text{maximize} & b^\mathsf{T}\tilde{y} \\ & \text{subject to} & \widetilde{\mathcal{A}}\widetilde{X} = b & & \text{subject to} & \widetilde{\mathcal{A}}^*\tilde{y} + \widetilde{S} = \widetilde{C} \\ & & \widetilde{X} \in \mathbb{S}^n_+ & & & \widetilde{S} \in \mathbb{S}^n_+ \end{array} \tag{15}$$

with primal variable $\widetilde{X} \in \mathbb{S}^n$ and dual variables $(\tilde{y}, \widetilde{S}) \in \mathbb{R}^m \times \mathbb{S}^n$. The coefficients are $\widetilde{C} := Q_\star^\mathsf{T} C Q_\star$ and $\widetilde{A}_i := Q_\star^\mathsf{T} A_i Q_\star$ for all $i \in [m]$. If we apply three-step ADMM (2) to the modified SDP (15), starting at $\widetilde{X}^{(0)} := Q_\star^\mathsf{T} X^{(0)} Q_\star$ and $\widetilde{S}^{(0)} := Q_\star^\mathsf{T} S^{(0)} Q_\star$, straightforward calculations show that the generated sequence $(\widetilde{X}^{(k)}, \widetilde{y}^{(k)}, \widetilde{S}^{(k)})$ is related to $(X^{(k)}, y^{(k)}, S^{(k)})$ as follows:

$$\widetilde{X}^{(k)} = Q_\star^\mathsf{T} X^{(k)} Q_\star, \qquad \tilde{y}^{(k)} = y^{(k)}, \qquad \widetilde{S}^{(k)} = Q_\star^\mathsf{T} S^{(k)} Q_\star, \quad \text{for all } k \in \mathbb{N}.$$

Moreover, the sequence $(\widetilde{X}^{(k)}, \widetilde{y}^{(k)}, \widetilde{S}^{(k)})$ converges to $(Q_\star^\mathsf{T} X_\star Q_\star, y_\star, Q_\star^\mathsf{T} S_\star Q_\star)$, a KKT point of the modified SDP (15).

4.2 Local Linearization

Now we study the local behavior of one-step ADMM (4). In particular, we linearize the ADMM iteration when near optimum. For ease of representation, we define

$$\mathcal{P} := \Pi_{\mathcal{R}(\mathcal{A}^*)} = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}, \;\; \mathcal{P}^\perp := \operatorname{Id} - \mathcal{P}, \;\; \Omega^\perp := E_n - \Omega, \;\; \Theta^\perp := E_{(n-r)\times r} - \Theta,$$

where recall E_r (resp., $E_{(n-r)\times r}$) is the all-ones matrix of size $n \times n$ (resp., $(n-r) \times r$). With the above abbreviations, we rewrite the iteration (4) as

$$Z^{(k+1)} - Z_\star = \mathcal{M}(Z^{(k)} - Z_\star) + \Psi^{(k)}, \tag{16}$$

where

$$\mathcal{M}(H) := \mathcal{P}(\Omega^\perp \circ H) + \mathcal{P}^\perp(\Omega \circ H), \tag{17}$$

$$\Psi^{(k)} := (\operatorname{Id} - 2\mathcal{P})(\Pi_{\mathbb{S}^n_+}(Z^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \Omega \circ (Z^{(k)} - Z_\star)). \tag{18}$$

This reformulation (16) says that when near optimum, the residual $H^{(k+1)} := Z^{(k+1)} - Z_\star$ is almost a linear transformation of the previous residual $H^{(k)} := Z^{(k)} - Z_\star$, with a quadratic correction term $\Psi^{(k)} = \mathcal{O}(\|H^{(k)}\|_2^2)$ (see Lemma 1).

Proof of (16). Note that $Z_\star$ is a fixed point of (4), *i.e.,*

$$Z_\star = \mathcal{P}(-2\Pi_{\mathbb{S}^n_+}(Z_\star) + Z_\star) + \Pi_{\mathbb{S}^n_+}(Z_\star) + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}b - \sigma(\operatorname{Id} - \mathcal{P})C.$$

Substituting back into (4) yields

$$\begin{aligned}
Z^{(k+1)} - Z_\star &= (\mathrm{Id} - 2\mathcal{P})(\Pi_{\mathbb{S}^n_+}(Z^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star)) + \mathcal{P}(Z^{(k)} - Z_\star) \\
&= (\mathrm{Id} - 2\mathcal{P})\Omega \circ (Z^{(k)} - Z_\star) + \mathcal{P}(Z^{(k)} - Z_\star) \\
&\quad + (\mathrm{Id} - 2\mathcal{P})(\Pi_{\mathbb{S}^n_+}(Z^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \Omega \circ (Z^{(k)} - Z_\star)) \\
&= (\mathrm{Id} - 2\mathcal{P})\Omega \circ (Z^{(k)} - Z_\star) + \mathcal{P}(Z^{(k)} - Z_\star) + \Psi^{(k)},
\end{aligned}$$

by the definition of $\Psi^{(k)}$. Moreover, for any $H \in \mathbb{S}^n$, we have

$$\begin{aligned}
(\mathrm{Id} - 2\mathcal{P})\Omega \circ H - \mathcal{P}H &= \mathcal{P}^\perp(\Omega \circ H) - \mathcal{P}(\Omega \circ H) + \mathcal{P}(H) \\
&= \mathcal{P}^\perp(\Omega \circ H) + \mathcal{P}(\Omega^\perp \circ H) \\
&= \mathcal{M}(H),
\end{aligned}$$

where the first line uses $\mathcal{P}^\perp = \mathrm{Id} - \mathcal{P}$ and the second line uses $\Omega^\perp = E_r - \Omega$. □

4.3 Properties of $\mathcal{M}$ and $\Psi^{(k)}$

Now we represent some properties of the linear operator $\mathcal{M}$ and the residual $\Psi^{(k)}$ that will be used to establish the linear convergence of ADMM. In particular, we characterize the nonempty set of fixed points $\mathrm{Fix}(\mathcal{M}) := \{H \mid \mathcal{M}(H) = H\}$; see Proposition 1. Throughout this subsection, we partition the matrix H (or $H^{(k)}$) as in (13) with $r := \mathrm{rank}(X_\star)$.

Lemma 2. *Suppose $\Omega \in \mathbb{S}^n$ is defined as in (11) with $r := \mathrm{rank}(X_\star)$. For any matrix $H \in \mathbb{S}^n$ as partitioned in (13), it holds that*

$$\left\langle \Omega \circ H, \Omega^\perp \circ H \right\rangle = 2\left\langle \Theta \circ H_O, \Theta^\perp \circ H_O \right\rangle \geq 0 \tag{19}$$

with equality only if $H_O = 0$, and that

$$\|H\|_\mathsf{F}^2 - \|\mathcal{M}(H)\|_\mathsf{F}^2 = \|\mathcal{P}(\Omega \circ H)\|_\mathsf{F}^2 + \|\mathcal{P}^\perp(\Omega^\perp \circ H)\|_\mathsf{F}^2 + 4\left\langle \Theta \circ H_O, \Theta^\perp \circ H_O \right\rangle. \tag{20}$$

Proof. From the definition of Ω (11) and the partition of H (13), we see that

$$\begin{aligned}
\left\langle \Omega \circ H, \Omega^\perp \circ H \right\rangle &= \left\langle \begin{bmatrix} E_r & \Theta^\mathsf{T} \\ \Theta & 0 \end{bmatrix} \circ \begin{bmatrix} H_X & H_O^\mathsf{T} \\ H_O & H_S \end{bmatrix}, \begin{bmatrix} 0 & (\Theta^\perp)^\mathsf{T} \\ \Theta^\perp & E_{n-r} \end{bmatrix} \circ \begin{bmatrix} H_X & H_O^\mathsf{T} \\ H_O & H_S \end{bmatrix} \right\rangle \\
&= \left\langle \begin{bmatrix} H_X & \Theta^\mathsf{T} \circ H_O^\mathsf{T} \\ \Theta \circ H_O & 0 \end{bmatrix}, \begin{bmatrix} 0 & (\Theta^\perp)^\mathsf{T} \circ H_O^\mathsf{T} \\ \Theta^\perp \circ H_O & H_S \end{bmatrix} \right\rangle \\
&= 2\left\langle \Theta \circ H_O, \Theta^\perp \circ H_O \right\rangle \geq 0.
\end{aligned} \tag{21}$$

Since all the entries in Θ and $\Theta^\perp$ are strictly positive, the inner product (21) is zero if and only if $H_O = 0$.

To show the second conclusion, we first decompose H as

$$H = \mathcal{P}(\Omega \circ H) + \mathcal{P}(\Omega^\perp \circ H) + \mathcal{P}^\perp(\Omega \circ H) + \mathcal{P}^\perp(\Omega^\perp \circ H). \tag{22}$$

Then we have

$$\begin{aligned}
\|H\|_\mathsf{F}^2 = {}& \|\mathcal{P}(\Omega \circ H)\|_\mathsf{F}^2 + \|\mathcal{P}(\Omega^\perp \circ H)\|_\mathsf{F}^2 + \|\mathcal{P}^\perp(\Omega \circ H)\|_\mathsf{F}^2 + \|\mathcal{P}^\perp(\Omega^\perp \circ H)\|_\mathsf{F}^2 \\
&+ 2\left\langle \mathcal{P}(\Omega \circ H), \mathcal{P}(\Omega^\perp \circ H) \right\rangle + 2\left\langle \mathcal{P}^\perp(\Omega \circ H), \mathcal{P}^\perp(\Omega^\perp \circ H) \right\rangle,
\end{aligned}$$

and

$$\|\mathcal{M}(H)\|_\mathsf{F}^2 = \|\mathcal{P}^\perp(\Omega \circ H) + \mathcal{P}(\Omega^\perp \circ H)\|_\mathsf{F}^2 = \|\mathcal{P}^\perp(\Omega \circ H)\|_\mathsf{F}^2 + \|\mathcal{P}(\Omega^\perp \circ H)\|_\mathsf{F}^2.$$

Combining both expressions with (19) gives the desirable result. □

Proposition 1. *The linear operator $\mathcal{M} : \mathbb{S}^n \to \mathbb{S}^n$ has the following properties.*

(a) $\mathcal{M}$ is firmly nonexpansive under the Frobenius norm.

(b) The sequence $\{\mathcal{M}^k\}_{k=1}^{\infty}$ converges to $\Pi_{\mathrm{Fix}(\mathcal{M})}$.

(c) $H \in \mathrm{Fix}(\mathcal{M})$ if and only if the following three conditions holds

$$H_O = 0, \qquad \begin{bmatrix} H_X & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{N}(\mathcal{A}), \qquad \begin{bmatrix} 0 & 0 \\ 0 & H_S \end{bmatrix} \in \mathcal{R}(\mathcal{A}^*). \tag{23}$$

(d) $\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}} < 1$.

Proof. Part (a): We verify the firm nonexpansiveness of $\mathcal{M}$ via its definition:

$$\begin{aligned}
&\langle \mathcal{M}(H), H \rangle \\
&= \left\langle \mathcal{P}(\Omega^\perp \circ H) + \mathcal{P}^\perp(\Omega \circ H), \mathcal{P}(\Omega \circ H) + \mathcal{P}(\Omega^\perp \circ H) + \mathcal{P}^\perp(\Omega \circ H) + \mathcal{P}^\perp(\Omega^\perp \circ H) \right\rangle \\
&= \|\mathcal{M}(H)\|_{\mathsf{F}}^2 + \left\langle \mathcal{P}(\Omega^\perp \circ H) + \mathcal{P}^\perp(\Omega \circ H), \mathcal{P}(\Omega \circ H) + \mathcal{P}^\perp(\Omega^\perp \circ H) \right\rangle \\
&= \|\mathcal{M}(H)\|_{\mathsf{F}}^2 + \left\langle \mathcal{P}(\Omega^\perp \circ H), \mathcal{P}(\Omega \circ H) \right\rangle + \left\langle \mathcal{P}^\perp(\Omega \circ H), \mathcal{P}^\perp(\Omega^\perp \circ H) \right\rangle \\
&= \|\mathcal{M}(H)\|_{\mathsf{F}}^2 + \left\langle \Omega^\perp \circ H, \Omega \circ H \right\rangle \qquad (24a) \\
&\geq \|\mathcal{M}(H)\|_{\mathsf{F}}^2, \qquad (24b)
\end{aligned}$$

where (24a) uses the fact that $\mathcal{P}\mathcal{P}^\perp = 0$ and (24b) uses (19).

Part (b) follows readily from part (a) and monotone operator theory; see [6, Proposition 5.16 (ii)] and [5, Corollary 2.7 (ii)].

Part (c): From the decomposition of H (22) and the definition of $\mathcal{M}$ (17), we see that

$$H \in \mathrm{Fix}(\mathcal{M}) \qquad \Longleftrightarrow \qquad \mathcal{P}^\perp(\Omega^\perp \circ H) = 0 \text{ and } \mathcal{P}(\Omega \circ H) = 0.$$

On one hand, if $H \in \mathrm{Fix}(\mathcal{M})$, we conclude from Lemma 2 that H_O has to be zero. Then expanding $\mathcal{P}(\Omega \circ H) = 0$ gives

$$\mathcal{P}\left(\begin{bmatrix} H_X & 0 \\ 0 & 0 \end{bmatrix} \right) = 0,$$

which is equivalent to the second condition in (23) (since $\mathcal{P} := \Pi_{\mathcal{N}(\mathcal{A})^\perp}$). Similarly, expanding $\mathcal{P}^\perp(\Omega^\perp \circ H) = 0$ gives the last condition in (23).

On the other hand, the first two conditions in (23) imply that $\mathcal{P}(\Omega \circ H) = 0$ (since all the entries in Θ zero strictly positive). Similarly, $H_O = 0$ and the last condition in (23) imply $\mathcal{P}^\perp(\Omega^\perp \circ H) = 0$. Combining the two results yields $H \in \mathrm{Fix}(\mathcal{M})$.

Part (d): is equivalent to show that $\|(\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})})H\|_{\mathsf{F}} < \|H\|_{\mathsf{F}}$ for any nonzero H. If $H \in \mathrm{Fix}(\mathcal{M})$ (and $H \neq 0$), then $\|(\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})})H\|_{\mathsf{F}} = 0 < \|H\|_{\mathsf{F}}$. Otherwise, $H \notin \mathrm{Fix}(\mathcal{M})$, and at least one of the three conditions in (23) is not satisfied. So, at least one of the three terms on the right-hand side of (20) is positive, which implies the desirable result. ∎

Proposition 2. *There exist two constants $\bar{k}_\Psi \in \mathbb{N}$ and $\alpha_\Psi > 0$ such that for any integer $k \geq \bar{k}_\Psi$, it holds that*

$$\|\Psi^{(k)}\|_{\mathsf{F}} \leq \alpha_\Psi \cdot \|H_O^{(k)}\|_{\mathsf{F}} \cdot \|H^{(k)}\|_{\mathsf{F}},$$

where $H^{(k)} := Z^{(k)} - Z_\star$ is partitioned as in (13).

Proof. The linear operator $2\mathcal{P} - \mathrm{Id}$ is the reflection operator, and thus preserves the Frobenius norm. Thus, we have from the definition of $\Psi^{(k)}$ (18) that

$$\begin{aligned}
\|\Psi^{(k)}\|_{\mathsf{F}} &= \|\Pi_{\mathbb{S}^n_+}(Z^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \Omega \circ (Z^{(k)} - Z_\star)\|_{\mathsf{F}} \\
&= \|\Pi_{\mathbb{S}^n_+}(Z_\star + H^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \Omega \circ H^{(k)}\|_{\mathsf{F}}.
\end{aligned}$$

The desirable result then follows from Theorem 2 and the fact that $H^{(k)} \to 0$ as $k \to \infty$. ∎

5 Local Linear Convergence with Nondegeneracy

In this section, we establish local linear convergence of ADMM when primal and dual nondegeneracy holds at optimum. In this case, the pair of SDPs (1) has a unique KKT point and Assumption 2 is equivalent to merely existence of a strictly complementary solution, which is a common regularity condition for SDP in the literature.

We start with the simple characteristic of $\mathrm{Fix}(\mathcal{M})$ when nondegeneracy holds.

Lemma 3. *Suppose Assumptions 1 and 2, primal nondegeneracy (8) and dual nondegeneracy (9) hold. Then, it holds that* $\mathrm{Fix}(\mathcal{M}) = \{0\}$.

Proof. For any $H \in \mathrm{Fix}(\mathcal{M})$, we see from Proposition 1 (c) and the definition of $\mathcal{T}_{X_\star}^{\perp}$ that

$$\begin{bmatrix} 0 & 0 \\ 0 & H_S \end{bmatrix} \in \mathcal{T}_{X_\star}^{\perp} \text{ and } \begin{bmatrix} 0 & 0 \\ 0 & H_S \end{bmatrix} \in \mathcal{R}(\mathcal{A}^*).$$

Yet primal nondegeneracy suggests $\mathcal{T}_{X_\star}^{\perp} \cap \mathcal{R}(\mathcal{A}^*) = \{0\}$. Thus, $H_S = 0$.

Similarly, from Proposition 1 (c) and the definition of $\mathcal{T}_{S_\star}^{\perp}$, we conclude that

$$\begin{bmatrix} H_X & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{T}_{S_\star}^{\perp} \text{ and } \begin{bmatrix} H_X & 0 \\ 0 & 0 \end{bmatrix} \in \mathcal{N}(\mathcal{A}).$$

Yet dual nondegeneracy suggests that $\mathcal{T}_{S_\star}^{\perp} \cap \mathcal{N}(\mathcal{A}) = \{0\}$. Thus, $H_X = 0$.

Therefore, any point $H \in \mathrm{Fix}(\mathcal{M})$ must satisfy $H = 0$; *i.e.*, $\mathrm{Fix}(\mathcal{M}) = \{0\}$. □

Theorem 3. *Suppose Assumptions 1 and 2, primal nondegeneracy (8) and dual nondegeneracy (9) hold. For any $\rho \in (\|\mathcal{M}\|_{\mathrm{op}}, 1)$, there exists $\bar{k}_{\mathrm{ND}} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}_{\mathrm{ND}}$, it holds that*

$$\|Z^{(k+1)} - Z_\star\|_{\mathsf{F}} \leq \rho \|Z^{(k)} - Z_\star\|_{\mathsf{F}}.$$

Proof. Since $\mathrm{Fix}(\mathcal{M}) = \{0\}$, we have $\Pi_{\mathrm{Fix}(\mathcal{M})} = 0$ and $\|\mathcal{M}\|_{\mathrm{op}} < 1$ from Proposition 1. Then Proposition 2 implies that there exists $\bar{k}_\Psi \in \mathbb{N}$ such that $\|\Psi^{(k)}\|_{\mathsf{F}} \leq \alpha_\Psi \|H_O^{(k)}\|_{\mathsf{F}} \|H^{(k)}\|_{\mathsf{F}}$ for any integer $k \geq \bar{k}_\Psi$. Then, convergence of ADMM suggests that for any $\rho \in (\|\mathcal{M}\|_{\mathrm{op}}, 1)$, there exists $\bar{k}_O$ such that for any integer $k \geq \max\{\bar{k}_\Psi, \bar{k}_O\} =: \bar{k}_{\mathrm{ND}}$, we have $\alpha_\Psi \|H_O^{(k)}\|_{\mathsf{F}} \leq \rho - \|\mathcal{M}\|_{\mathrm{op}}$ and

$$\|\Psi^{(k)}\|_{\mathsf{F}} \leq (\rho - \|\mathcal{M}\|_{\mathrm{op}}) \cdot \|H^{(k)}\|_{\mathsf{F}} = (\rho - \|\mathcal{M}\|_{\mathrm{op}}) \cdot \|Z^{(k)} - Z_\star\|_{\mathsf{F}}.$$

Finally,

$$\begin{aligned} \|Z^{(k+1)} - Z_\star\|_{\mathsf{F}} &= \|\mathcal{M}(Z^{(k)} - Z_\star) + \Psi^{(k)}\|_{\mathsf{F}} \\ &\leq \|\mathcal{M}\|_{\mathrm{op}} \cdot \|Z^{(k)} - Z_\star\|_{\mathsf{F}} + \|\Psi^{(k)}\|_{\mathsf{F}} \\ &\leq (\|\mathcal{M}\|_{\mathrm{op}} + \rho - \|\mathcal{M}\|_{\mathrm{op}}) \cdot \|Z^{(k)} - Z_\star\|_{\mathsf{F}} \\ &= \rho \|Z^{(k)} - Z_\star\|_{\mathsf{F}}. \end{aligned}$$

□

Remark 4. *Our proof framework can also cover the case where ND holds and SC fails. To stay consistent with our SC assumption, the detailed proof of this case is deferred to Appendix B. So, combining Theorem 3 and the results in Appendix B, we establish local linear convergence of ADMM for SDP under only the nondegeneracy conditions. Though this conclusion can be drawn from [22], our proof techniques are completely different from theirs and do not involve the metric subregularity of the KKT operator. Moreover, numerical evidence is provided in Section 8 to support the theoretical findings in Appendix B.*

6 Local R-linear Convergence without Nondegeneracy

Without two-side nondegeneracy, $\mathrm{Fix}(\mathcal{M})$ is not $\{0\}$ and the proof technique in Section 5 does not apply anymore. Another nice property of $\mathrm{Fix}(\mathcal{M})$ in Proposition 1 turns out to be useful: $\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}} < 1$. Specifically, this property motivates us to study the "projected sequence":

$$\begin{aligned}(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k+1)} &= (\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\mathcal{M}H^{(k)} + (\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)} \\ &= (\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})})(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)} + (\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)},\end{aligned}$$

where the last equality follows from

$$(\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})})(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})}) = \mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})} = (\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\mathcal{M}.$$

So, combining with the structure of $\mathrm{Fix}(\mathcal{M})$ and our refined error bound in Theorem 2, we are able to establish the (R-)linear convergence of several "partial" sequences

$$(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}, \qquad H_O^{(k)}, \qquad \Pi_{\mathcal{T}_{S_\star}}(X^{(k)}), \qquad \Pi_{\mathcal{T}_{X_\star}}(S^{(k)}),$$

where the last two terms correspond to the part of $X^{(k)}$ (resp., $S^{(k)}$) that lies outside the minimal face of $X_\star$ (resp., $S^{(k)}$); see Lemmas 5 and 6. (One may already notice that in the nondegenerate case where $\mathrm{Fix}(\mathcal{M}) = \{0\}$, the sequence $(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}$ is exactly $H^{(k)}$, and the proof is done at this step.)

So, what is missing in the more general, possibly degenerate case? It turns out that an error bound for $\Pi_{\mathbb{S}^n_+}$ alone is insufficient; a growth condition is needed that accounts for both the PSD cone and an affine set. More specifically, consider the spectrahedron $\mathcal{V} \cap \mathbb{S}^n_+$, where $\mathcal{V}$ is an affine space in $\mathbb{S}^n$. Following the convention in [53], we call $\mathrm{dist}(X, \mathcal{V} \cap \mathbb{S}^n_+)$ the *forward error* and $\mathrm{dist}(X, \mathcal{V}) + [-\lambda_{\min}(X)]_+$ the *backward error*. In the polyhderal case (*i.e.,* $\mathbb{S}^n_+$ reduces to the nonnegative orthant), the backward error and the forward error are in the same order [26], which leads to the *sharpness* condition and linear convergence of first-order methods in linear programming [3]. In the spectrahedron case, however, it is shown in [53] that

$$\text{forward error} = \mathcal{O}((\text{backward error})^{1/2})$$

under mild conditions. So SDPs are not sharp in general and linear convergence does not follow in a straightforward manner.

Fortunately, by exploiting the geometry of the PSD cone, it is shown in [53, Lemma 2.3] that the forward error is in the same order as the backward error with respect to the regularized system

$$\mathcal{V} \cap \mathrm{minface}\left(X_\star, \mathbb{S}^n_+\right), \quad \text{where } \mathrm{minface}\left(X_\star, \mathbb{S}^n_+\right) := \left\{ \begin{bmatrix} \Gamma & 0 \\ 0 & 0 \end{bmatrix} \,\middle|\, \Gamma \in \mathbb{S}^r_+ \right\} = \mathbb{S}^n_+ \cap \mathcal{T}_{S_\star}^{\perp}.$$

(The simple characteristic of the minimal face needs the assumption, made without loss of generality, that $X_\star$ is diagonal.) Extending the conclusion in [53, Lemma 2.3], we obtain a linear growth condition on the distance to optimality. More specifically, we upper bound the distance from $Z^{(k)}$ to the optimal set by the sum of the following three terms:

$$\|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}, \qquad \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}}, \qquad \|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}},$$

where the last two terms correspond to the part of $X^{(k)}$ (resp., $S^{(k)}$) that lies outside the minimal face of $X_\star$ (resp., $S_\star$); see Lemmas 7 and 8. Finally, combining the two ingredients (convergence of some partial sequences and the new growth condition) yields the desirable linear convergence guarantees of ADMM without nondegeneracy conditions.

Below we dive into the details, we remind that without two-side nondegeneracy, the primal and dual solutions may not be unique. So we denote by $\mathcal{X}_\star$ the optimal set for primal SDP in (1), by $\mathcal{S}_\star$ the set of dual optimal S, and by $\mathcal{Z}_\star$ the set of fixed points for the one-step ADMM (4).

We begin our analysis with some basic results on ADMM, of which the proof mainly uses the ADMM update rule (4). In fact, the inequality (26) is a special case of [67, Proposition 3.1].

Lemma 4. *The sequence* $\{Z^{(k)}\}$ *generated by one-step ADMM* (4) *satisfies*

$$\|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 = \|\mathcal{P}(X^{(k)} - \widetilde{X})\|_{\mathsf{F}}^2 + \sigma^2 \|\mathcal{P}^{\perp}(S^{(k)} - C)\|_{\mathsf{F}}^2, \tag{25}$$

where $\widetilde{X}$ *is an arbitrary matrix satisfying* $\mathcal{A}\widetilde{X} = b$*. And*

$$\|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 \le \operatorname{dist}^2(Z^{(k)}, \mathcal{Z}_\star) - \operatorname{dist}^2(Z^{(k+1)}, \mathcal{Z}_\star), \tag{26}$$

for all $k \in \mathbb{N}$.

Proof. See Appendix C.1. □

6.1 R-linear Decay outside Minimal Faces

Lemma 5. *Suppose Assumptions 1 and 2 hold. Let* $H^{(k)} := Z^{(k)} - Z_\star$, $k \in \mathbb{N}$, *be partitioned as in* (13)*. Then, for any* $\rho \in (\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}}, 1)$*, there exists* $\bar{k}_\rho \in \mathbb{N}$ *such that for any integer* $k \ge \bar{k}_\rho$*, it holds that*

$$\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k+1)}\|_{\mathsf{F}} \le \rho \|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}.$$

Moreover, $\|H_O^{(k)}\|_{\mathsf{F}}$ *converges R-linearly.*

Proof. We first show that

$$\frac{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)}\|_{\mathsf{F}}}{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}} \to 0 \quad \text{as } k \to \infty.$$

To see this, we first note from Proposition 1 (c) that the off-block-diagonal part of $\Pi_{\mathrm{Fix}(\mathcal{M})}(H^{(k)})$ is zero, which implies that $(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}$ and $H^{(k)}$ have the same off-block-diagonal part. So,

$$\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}} \ge \sqrt{2}\|H_O^{(k)}\|_{\mathsf{F}}.$$

Then, we conclude from Proposition 2 that there exist $\bar{k}_\Psi \in \mathbb{N}$ and $\alpha_\Psi > 0$ such that for any integer $k \ge \bar{k}_\Psi$, it holds that

$$\frac{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)}\|_{\mathsf{F}}}{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}} \le \frac{\|\Psi^{(k)}\|_{\mathsf{F}}}{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}} \le \frac{\alpha_\Psi \|H_O^{(k)}\|_{\mathsf{F}} \|H^{(k)}\|_{\mathsf{F}}}{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}} \le \frac{\alpha_\Psi}{\sqrt{2}} \|H^{(k)}\|_{\mathsf{F}},$$

which goes to 0 as $k \to \infty$.

Finally, the R-linear convergence of $\|H_O^{(k)}\|_{\mathsf{F}}$ follows naturally from the fact that $\|H_O^{(k)}\|_{\mathsf{F}} \le \|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}$. □

Theorem 2 plays a vital role in the proof of Lemma 5. If we used Lemma 1, we could only upper bound $\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)}\|_{\mathsf{F}}$ by $\|H^{(k)}\|_{\mathsf{F}}^2$. This could only imply

$$\frac{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})\Psi^{(k)}\|_{\mathsf{F}}}{\|(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}\|_{\mathsf{F}}} = \mathcal{O}\left(\frac{\|H^{(k)}\|_{\mathsf{F}}^2}{\|H_O^{(k)}\|_{\mathsf{F}}}\right). \tag{27}$$

In [35], the authors investigate general ADMM for convex problems with partially smooth objectives and attempt to establish the linear convergence of the projected sequence $(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{M})})H^{(k)}$. In [35, pp. 911, line 5], the authors assert (without providing a justification) that the left-hand side of (27) vanishes as $k \to \infty$. With the refined error bound in Theorem 2, our analysis confirms this claim in the context of SDP. However, its validity in the more general convex setting remains unclear to us.

Remark 5. *The R-linear convergence of* $\|H_O\|_{\mathsf{F}}$ *requires the assumption, made without loss of generality, that* $X_\star$ *and* $S_\star$ *are diagonal. Otherwise, when* $Q_\star \ne I$*, the R-linearly convergent sequence is* $\|Q_{\star,S}^{\mathsf{T}} H Q_{\star,X}\|_{\mathsf{F}}$*, where* $Q_\star = \begin{bmatrix} Q_{\star,X} & Q_{\star,S} \end{bmatrix}$.

Lemma 6. *Suppose Assumptions 1 and 2 hold. for any $\rho \in (\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}}, 1)$, there exists $\bar{k}_{\mathcal{T}} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}_{\mathcal{T}}$, the two norms*

$$\|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} \quad \textit{and} \quad \|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}}$$

converge R-linearly.

Proof. For any $k \in \mathbb{N}$, we have

$$\begin{aligned}
\|\Psi^{(k)}\|_{\mathsf{F}} &= \|\Pi_{\mathbb{S}^n_+}(Z_\star + H^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \Omega \circ H^{(k)}\|_{\mathsf{F}} \\
&= \|X^{(k)} - X_\star - \Omega \circ H^{(k)}\|_{\mathsf{F}} \qquad (28a)\\
&\geq \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)} - X_\star - \Omega \circ H^{(k)})\|_{\mathsf{F}}. \qquad (28b)
\end{aligned}$$

where (28a) uses $X^{(k)} = \Pi_{\mathbb{S}^n_+}(Z^{(k)})$. Then, from the definition of $\mathcal{T}_{S_\star}$ (7), we have

$$\Pi_{\mathcal{T}_{S_\star}}(\Omega \circ H^{(k)}) = \begin{bmatrix} 0 & \Theta^\mathsf{T} \circ H_O^\mathsf{T} \\ \Theta \circ H_O & 0 \end{bmatrix}. \qquad (29)$$

Combining (28b), and (29) and Proposition 2, we conclude that there exists $\bar{k}_\Psi$ such that for any integer $k \geq \bar{k}_\Psi$, we have

$$\begin{aligned}
\|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} &= \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)} - X_\star)\|_{\mathsf{F}} \\
&\leq \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)} - X_\star - \Omega \circ H^{(k)})\|_{\mathsf{F}} + \|\Pi_{\mathcal{T}_{S_\star}}(\Omega \circ H^{(k)})\|_{\mathsf{F}} \\
&\leq \|\Psi^{(k)}\|_{\mathsf{F}} + \sqrt{2}\|H_O^{(k)}\|_{\mathsf{F}} \\
&\leq (\alpha_\Psi \|H^{(k)}\|_{\mathsf{F}} + \sqrt{2})\|H_O^{(k)}\|_{\mathsf{F}},
\end{aligned}$$

The convergence of ADMM suggests that for sufficiently large $k \in \mathbb{N}$, the residual $H^{(k)}$ is bounded, and thus $\|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)} - X_\star)\|_{\mathsf{F}}$ is bounded above by a multiple of $\|H_O^{(k)}\|_{\mathsf{F}}$, a R-linearly convergent sequence (Lemma 5).

The second part of the lemma follows similarly since $S^{(k)} = (1/\sigma)\Pi_{\mathbb{S}^n_+}(-Z^{(k)})$. ☐

6.2 Linear Growth of Distance to Optimality

In this section, we present the one-iteration analysis for our convergence measure $\mathrm{dist}(Z^{(k)}, \mathcal{Z}_\star)$. The following lemma is inspired by [53, Lemma 2.3] and gives the regularized backward error for the (scaled) KKT system.

Lemma 7. *Let $(X_\star, y_\star, S_\star)$ be the convergent point of ADMM (2) satisfying strict complementarity (10). Then, there exist three positive constants $(\delta_X, \delta_S, \kappa)$ such that for all $(X, S) \in \mathbb{S}^n \times \mathbb{S}^n$ with $\|X\|_{\mathsf{F}} \leq \delta_X$ and $\|\sigma S\|_{\mathsf{F}} \leq \delta_S$, it holds that*

$$\begin{aligned}
&\kappa \cdot \mathrm{dist}((X, \sigma S), \mathcal{X}_\star \times (\sigma \mathcal{S}_\star)) \\
&\leq \|\mathcal{P}(X - \widetilde{X})\|_{\mathsf{F}} + \|\mathcal{P}^\perp(\sigma S - \sigma C)\|_{\mathsf{F}} + |\langle X, \sigma C\rangle + \langle \widetilde{X}, \sigma S\rangle - \langle \widetilde{X}, \sigma C\rangle| \\
&\quad + [-\lambda_{\min}(X)]_+ + [-\lambda_{\min}(\sigma S)]_+ + \|\Pi_{\mathcal{T}_{S_\star}}(X)\|_{\mathsf{F}} + \|\Pi_{\mathcal{T}_{X_\star}}(\sigma S)\|_{\mathsf{F}},
\end{aligned}$$

where $\widetilde{X}$ is an arbitrary matrix with $\mathcal{A}\widetilde{X} = b$ and $\sigma > 0$ is the parameter in ADMM.

Proof. See Appendix C.2. ☐

Lemma 8. *Suppose Assumptions 1 and 2 hold. Then, there exists $\bar{k}_Z \in \mathbb{N}$ and $\alpha_Z > 0$ such that for any integer $k \geq \bar{k}_Z$, it holds that*

$$\mathrm{dist}(Z^{(k)}, \mathcal{Z}_\star) \leq \alpha_Z(\|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}} + \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} + \sigma\|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}}),$$

Proof. We first bound the distance $\operatorname{dist}(Z^{(k)}, \mathcal{Z}_\star)$ by the distance from $(X^{(k)}, \sigma S^{(k)})$ to the set $\mathcal{X}_\star \times (\sigma \mathcal{S}_\star)$ as follows:

$$\begin{aligned}
\operatorname{dist}(Z^{(k)}, \mathcal{Z}_\star) &= \inf\{\|Z^{(k)} - Z\|_{\mathsf{F}} \mid Z \in \mathcal{Z}_\star\} \\
&= \inf\{\|Z^{(k)} - (X - \sigma S)\|_{\mathsf{F}} \mid X \in \mathcal{X}_\star, S \in \mathcal{S}_\star\} \\
&= \inf\{\|(\widetilde{X} - \sigma \widetilde{S}) - (X - \sigma S)\|_{\mathsf{F}} \mid \widetilde{X} - \sigma \widetilde{S} = Z^{(k)}, X \in \mathcal{X}_\star, S \in \mathcal{S}_\star\} \\
&\leq \sqrt{2} \cdot \inf\left\{ \sqrt{\|\widetilde{X} - X\|_{\mathsf{F}}^2 + \sigma^2 \|\widetilde{S} - S\|_{\mathsf{F}}^2} \;\middle|\; \widetilde{X} - \sigma \widetilde{S} = Z^{(k)}, X \in \mathcal{X}_\star, S \in \mathcal{S}_\star \right\} \\
&\leq \sqrt{2} \cdot \operatorname{dist}((X^{(k)}, \sigma S^{(k)}), \mathcal{X}_\star \times (\sigma \mathcal{S}_\star)).
\end{aligned}$$

Then, we bound $\operatorname{dist}((X^{(k)}, \sigma S^{(k)}), \mathcal{X}_\star \times (\sigma \mathcal{S}_\star))$ using Lemma 7 and the facts that $X^{(k)} \in \mathbb{S}_+^n$ and $S^{(k)} \in \mathbb{S}_+^n$ for all $k \in \mathbb{N}$. More specifically, there exist $(\bar{k}_Z, \delta_X, \delta_S) \in \mathbb{N} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ such that for any integer $k \geq \bar{k}_Z$, we have $\|X^{(k)}\|_{\mathsf{F}} \leq \delta_X$, $\|S^{(k)}\|_{\mathsf{F}} \leq \delta_S$ and

$$\begin{aligned}
&\kappa \cdot \operatorname{dist}((X^{(k)}, \sigma S^{(k)}), \mathcal{X}_\star \times (\sigma \mathcal{S}_\star)) \\
&\leq \|\mathcal{P}(X^{(k)} - \widetilde{X})\|_{\mathsf{F}} + \sigma \|\mathcal{P}^\perp(S^{(k)} - C)\|_{\mathsf{F}} + \sigma |\langle X^{(k)}, C\rangle + \langle \widetilde{X}, S^{(k)}\rangle - \langle \widetilde{X}, C\rangle| \\
&\quad + \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} + \sigma \|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}},
\end{aligned} \tag{30}$$

where $\widetilde{X}$ is an arbitrary matrix satisfying $\mathcal{A}\widetilde{X} = b$. The inner product on the right-hand side of (30) can be further bounded by

$$\begin{aligned}
&\sigma |\langle X^{(k)}, C\rangle + \langle \widetilde{X}, S^{(k)}\rangle - \langle \widetilde{X}, C\rangle| \\
&= \sigma |\langle X^{(k)}, C\rangle + \langle \widetilde{X}, S^{(k)}\rangle - \langle \widetilde{X}, C\rangle| \\
&= \sigma |\langle X - \widetilde{X}, S^{(k)} - C\rangle| && \text{(31a)} \\
&= \sigma |\langle \mathcal{P}(X - \widetilde{X}), \mathcal{P}(S^{(k)} - C)\rangle + \langle \mathcal{P}^\perp(X - \widetilde{X}), \mathcal{P}^\perp(S^{(k)} - C)\rangle| \\
&\leq \sigma \|\mathcal{P}(X^{(k)} - \widetilde{X})\|_{\mathsf{F}} \|\mathcal{P}(S^{(k)} - C)\|_{\mathsf{F}} + \sigma \|\mathcal{P}^\perp(X^{(k)} - \widetilde{X})\|_{\mathsf{F}} \|\mathcal{P}^\perp(S^{(k)} - C)\|_{\mathsf{F}} \\
&\leq \alpha_Z' (\|\mathcal{P}(X^{(k)} - \widetilde{X})\|_{\mathsf{F}} + \sigma \|\mathcal{P}^\perp(S^{(k)} - C)\|_{\mathsf{F}}), && \text{(31b)}
\end{aligned}$$

where (31a) uses the fact $\langle X^{(k)}, S^{(k)}\rangle = 0$ and in (31b) we define

$$\alpha_Z' := \max\{\sigma(\delta_S + \|C\|_{\mathsf{F}}),\, \delta_X + \|\mathcal{A}^\dagger b\|_{\mathsf{F}}\}$$

(recall $\mathcal{A}^\dagger b$ is a valid choice for $\widetilde{X}$). Finally, combining (30), (31b) with (25) in Lemma 4 and denoting $\alpha_Z := \sqrt{2}(1 + \alpha_Z')/\kappa$ give the desirable result. $\square$

6.3 Main Theorem

Now we are ready to present our main theorem.

Theorem 4. *Suppose Assumptions 1 and 2 hold. Then, for any $\rho_0 \in (\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}}, 1)$, there exists $\bar{k} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}$, the distance to optimality $\operatorname{dist}(Z^{(k)}, \mathcal{Z}_\star)$ converges R-linearly;* i.e., *there exists $(\alpha, \rho) \in \mathbb{R}_{++} \times (0, 1)$ such that*

$$\operatorname{dist}(Z^{(k)}, \mathcal{Z}_\star) \leq \alpha \rho^k.$$

Proof. Define $a^{(k)} := \operatorname{dist}(Z^{(k)}, \mathcal{Z}_\star)$ for brevity. From Lemma 6, we deduce that there exist $(\bar{k}_{\mathcal{T}}, \alpha_X, \alpha_S) \in \mathbb{N} \times \mathbb{R}_{++} \times \mathbb{R}_{++}$ such that for any integer $k \geq \bar{k}_{\mathcal{T}}$, we have

$$\|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} \leq \alpha_X \rho_0^k, \qquad \|\Pi_{\mathcal{T}_{S_\star}}(S^{(k)})\|_{\mathsf{F}} \leq \alpha_S \rho_0^k. \tag{32}$$

Then, with $\bar{k}_Z \in \mathbb{N}$ as defined in Lemma 8, we have for any integer $k \geq \bar{k} := \max\{\bar{k}_{\mathcal{T}}, \bar{k}_Z\} + 1$ that

$$\begin{aligned} a^{(k+1)} &\leq a^{(k)} && \text{(33a)} \\ &\leq \alpha_Z(\|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}} + \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} + \sigma\|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}}) && \text{(33b)} \\ &\leq \alpha_Z\Big(\sqrt{(a^{(k)})^2 - (a^{(k+1)})^2} + \|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} + \sigma\|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}}\Big) && \text{(33c)} \\ &\leq \alpha_Z\sqrt{(a^{(k)})^2 - (a^{(k+1)})^2} + \alpha_Z(\alpha_X + \alpha_S)\rho_0^k. && \text{(33d)} \end{aligned}$$

In (33a) we use (26) in Lemma 4, (33b) uses Lemma 8, (33c) uses (26) again, and finally (33d) uses (32).

Then, we partition the index set $\{k \in \mathbb{N} \mid k \geq \bar{k}\}$ into

$$\mathcal{I} := \{k \in \mathbb{N} \mid k \geq \bar{k},\, a^{(k+1)} \geq 2\alpha_Z(\alpha_X + \alpha_S)\rho_0^k\}$$

and its complement $\mathcal{I}^{\mathrm{c}} := \{k \in \mathbb{N} \mid k \geq \bar{k}\} \setminus \mathcal{I}$.

1. If $k \in \mathcal{I}$, then from (33d) we have

$$(a^{(k)})^2 - (a^{(k+1)})^2 \geq (a^{(k+1)} - \alpha_Z(\alpha_X + \alpha_S)\rho_0^2)^2 \geq \frac{1}{4\alpha_Z^2}(a^{(k+1)})^2,$$

   which implies that

$$a^{(k+1)} \leq \sqrt{\frac{4\alpha_Z^2}{1 + 4\alpha_Z^2}}\, a^{(k)}.$$

2. If $k \in \mathcal{I}^{\mathrm{c}}$, then readily we have

$$a^{(k+1)} \leq 2\alpha_Z(\alpha_X + \alpha_S)\rho_0^k.$$

Combining the two cases yields the desirable local R-linear convergence of $a^{(k)}$. More specifically, define

$$b^{(k)} := \max\{a^{(k)}, 2\alpha_Z(\alpha_X + \alpha_S)\rho_0^{k-1}\}, \qquad \rho := \max\left\{\sqrt{\frac{4\alpha_Z^2}{1 + 4\alpha_Z^2}}, \rho_0\right\} \in (0, 1),$$

and consider any pair $(k-1, k)$.

1. If $(k-1, k) \in \mathcal{I} \times \mathcal{I}$, then $b^{(k+1)} = a^{(k+1)} \leq \rho a^{(k)} = \rho b^{(k)}$.
2. If $(k-1, k) \in \mathcal{I} \times \mathcal{I}^{\mathrm{c}}$, then $b^{(k+1)} \leq 2\alpha_Z(\alpha_X + \alpha_S)\rho_0^k \leq \rho_0 a^{(k)} \leq \rho b^{(k)}$.
3. If $(k-1, k) \in \mathcal{I}^{\mathrm{c}} \times \mathcal{I}$, then $b^{(k+1)} = a^{(k+1)} \leq \rho a^{(k)} \leq \rho b^{(k)}$.
4. If $(k-1, k) \in \mathcal{I}^{\mathrm{c}} \times \mathcal{I}^{\mathrm{c}}$, then $b^{(k+1)} = 2\alpha_Z(\alpha_X + \alpha_S)\rho_0^k = \rho b^{(k)}$.

To conclude, $\{b^{(k)}\}$ is a linearly convergent sequence with rate $\rho \in (0, 1)$ and an upper bound for $\{a^{(k)}\}$. So, $a^{(k)} := \mathrm{dist}(Z^{(k)}, \mathcal{Z}_\star)$ converges R-linearly for sufficiently large $k \in \mathbb{N}$. □

7 Proof of the Refined Error Bound

In this section, we detail the proof of Theorem 2, which builds an error bound for the PSD cone projection:

$$\|\Pi_{\mathbb{S}_+^n}(Z + H) - \Pi_{\mathbb{S}_+^n}(Z) - \Omega \circ H\|_2 \leq \alpha_{\mathrm{EB}} \cdot \|H_O\|_2 \cdot \|H\|_2,$$

where we have assumed without loss of generality that Z is diagonal. A traditional way to compute the orthogonal projection onto the PSD cone is via eigenvalue decomposition. Though conceptually simple, this method destroys the block structure of $Z + H$ (as H is not diagonal). So, to prove Theorem 2, we advocate

Algorithm 1: An iterative elimination procedure for PSD cone projection

Input: A nonsingular matrix $Z \in \mathbb{S}^n$ with r positive eigenvalues, and a perturbation $H \in \mathbb{S}^n$.
Output: $\Pi_{\mathbb{S}^n_+}(Z+H) \leftarrow V_\infty$.

Initialization: $Z_0 := \begin{bmatrix} Z_{X,0} & Z_{O,0}^\top \\ Z_{O,0} & Z_{S,0} \end{bmatrix} \leftarrow Z + H$ and $Q_0 \leftarrow I_n$.

for $\ell = 0$ *to* ∞ **do**

(1) Solve the following Sylvester equation for $W_O \in \mathbb{R}^{(n-r)\times r}$

$$W_O Z_{X,\ell} + (-Z_{S,\ell}) W_O = Z_{O,\ell} \tag{36}$$

and obtain $W_{O,\ell} \leftarrow W_O$.

(2) Compute

$$W_\ell \leftarrow \begin{bmatrix} 0 & -W_{O,\ell}^\top \\ W_{O,\ell} & 0 \end{bmatrix}$$

$$Z_{\ell+1} := \begin{bmatrix} Z_{X,\ell+1} & Z_{O,\ell+1}^\top \\ Z_{O,\ell+1} & Z_{S,\ell+1} \end{bmatrix} \leftarrow \exp(W_\ell)^\top \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\top \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} \exp(W_\ell). \tag{37}$$

(3) Compute $Q_{\ell+1} \leftarrow Q_\ell \exp(W_\ell)$ and

$$V_{\ell+1} \leftarrow Q_{\ell+1} \begin{bmatrix} Z_{X,\ell+1} & 0 \\ 0 & 0 \end{bmatrix} Q_{\ell+1}^\top. \tag{38}$$

end

for a seemingly much more complex procedure inspired by iterative methods for eigenvalue decomposition (see, *e.g.,* [47]). We detail the iterative algorithm in Algorithm 1 and briefly discuss the high-level intuition here. Consider the matrix

$$Z + H = \operatorname{diag}(\lambda_1, \ldots, \lambda_r, \lambda_{r+1}, \ldots, \lambda_n) + \begin{bmatrix} H_X & H_O^\top \\ H_O & H_S \end{bmatrix} =: \begin{bmatrix} Z_{X,0} & Z_{O,0}^\top \\ Z_{O,0} & Z_{S,0} \end{bmatrix}. \tag{34}$$

When $\|H\|_2$ is sufficiently small, we have $Z_{X,0} = \Lambda_X + H_X \in \mathbb{S}^r_{++}$ and $Z_{S,0} = \Lambda_S + H_S \in \mathbb{S}^{n-r}_{--}$. Algorithm 1 explicitly constructs an orthogonal matrix Q_∞ that is close to I_n and satisfies

$$Q_\infty^\top \begin{bmatrix} Z_{X,0} & Z_{O,0}^\top \\ Z_{O,0} & Z_{S,0} \end{bmatrix} Q_\infty = \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & Z_{S,\infty} \end{bmatrix}. \tag{35}$$

(So roughly speaking, $Z_{X,\infty} \in \mathbb{S}^r_{++}$ (resp., $Z_{S,\infty} \in \mathbb{S}^{n-r}_{--}$) is also close to $Z_{X,0}$ (resp., $Z_{S,0}$).) To compute the orthogonal matrix Q_∞ in (35), we solve a series of Sylvester equations (36) for W_ℓ and show that the recursively defined matrix $Q_{\ell+1} \leftarrow Q_\ell \exp(W_\ell)$ converges to Q_∞. As we will see later, each Sylvester equation helps build a skew-symmetric matrix W_ℓ such that the off-block-diagonal part of $(I_n + W_\ell)^\top Z_\ell (I_n + W_\ell)$ is gradually removed at each iteration; see (37). Then, with Q_∞ computed (or approximated), we can derive a fine-grained error bound for

$$\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H = Q_\infty \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} Q_\infty^\top - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H,$$

which further leads to the conclusion in Theorem 2.

Proof outline. The proof of Theorem 2 is accomplished by exploiting the properties of the sequences generated by Algorithm 1.

1. We show that at each iteration, the Sylvester equation (36) is well-defined and has a unique solution. We ensure this by showing $Z_{X,\ell} \in \mathbb{S}^r_{++}, Z_{S,\ell} \in \mathbb{S}^{n-r}_{--}$ for all $\ell \in \mathbb{N}$.

2. We show that the limit of the sequence $\{V_\ell\}_{\ell=0}^\infty$ exists and is exactly $\Pi_{\mathbb{S}^n_+}(Z+H)$. This is achieved by showing the exponential decay of the three sequences

$$\|Z_{X,\ell+1} - Z_{X,\ell}\|_2, \qquad \|Z_{S,\ell+1} - Z_{S,\ell}\|_2, \qquad \|Q_{\ell+1} - Q_\ell\|_2.$$

3. Last, we show that

$$\begin{aligned}
&\|\Pi_{\mathbb{S}^n_+}(Z+H) - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H\|_2 \\
=\ & \left\| Q_\infty \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} Q_\infty^\mathsf{T} - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H \right\|_2 \\
\le\ & \left\| (I_n + W_0) \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} (I_n + W_0)^\mathsf{T} - \Pi_{\mathbb{S}^n_+}(Z) - \Omega \circ H \right\|_2 \\
&+ \left\| Q_\infty \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} Q_\infty^\mathsf{T} - (I_n + W_0) \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} (I_n + W_0)^\mathsf{T} \right\|_2
\end{aligned} \tag{39}$$

   Then, we bound the growth of the first term on the right-hand side of (39) by $\mathcal{O}(\|H_O\|_2 \cdot \|H\|_2)$ and that of the second term by $\mathcal{O}\left(\|H_O\|_2^2\right)$.

Remark 6. *We reiterate that in (34), we have assumed without loss of generality that Z is diagonal. Extension of the presented proof to the non-diagonal case is straightforward and detailed in Section 7.4.*

7.1 Step 1: Error Bound for Sylvester Equations

Lemma 9 shows that $Z_{X,\ell} \in \mathbb{S}^r_{++}$ and $Z_{S,\ell} \in \mathbb{S}^{n-r}_{--}$ imply the well-posedness of the Sylvester equation (36). Then, Lemma 10 proves that the definitenss of $Z_{X,\ell}$ and $Z_{S,\ell}$ holds as long as $Z_{X,0} \in \mathbb{S}^r_{++}$, $Z_{S,0} \in \mathbb{S}^{n-r}_{--}$ and $\|Z_{O,0}\|_2$ is sufficiently small. With these two lemmas, we complete Step 1 in the proof outline. For ease of notation, we define $d := \sqrt{\min\{r, n-r\}}$ and

$$\eta_\ell := \frac{d}{\lambda_{\min}(Z_{X,\ell}) - \lambda_{\max}(Z_{S,\ell})} \quad \text{for } \ell \in \mathbb{N}. \tag{40}$$

Lemma 9. *At iteration ℓ in Algorithm 1, suppose $Z_{X,\ell} \in \mathbb{S}^r_{++}$, $Z_{S,\ell} \in \mathbb{S}^{n-r}_{--}$, and $\|Z_{O,\ell}\|_2 \le \frac{3}{4\eta_\ell}$. Then, the Sylvester equation (36) has a unique solution $W_{O,\ell}$ satisfying $\|W_{O,\ell}\|_2 \le \eta_\ell \cdot \|Z_{O,\ell}\|_2$. Moreover, it holds that*

$$\begin{aligned}
&\max\{\|Z_{X,\ell+1} - Z_{X,\ell}\|_2, \|Z_{S,\ell+1} - Z_{S,\ell}\|_2, \|Z_{O,\ell+1}\|_2\} \\
\le\ & \left(\frac{4}{9}\eta_\ell^4 \cdot \|Z_0\|_2^3 + \frac{4}{3}\eta_\ell^3 \cdot \|Z_0\|_2^2 + \frac{13}{3}\eta_\ell^2 \cdot \|Z_0\|_2 + 4\eta_\ell\right) \cdot \|Z_{O,\ell}\|_2^2
\end{aligned}$$

for all $\ell \in \mathbb{N}$.

Proof. See Appendix D.1. □

The proof of Lemma 10 needs two auxiliary functions. Define $f(x) : [0, \infty) \mapsto \mathbb{R}$ as any fixed continuous and monotonically increasing function satisfying: (1) $f(0) = 0$; (2) $f(x) \ge \frac{9}{4}x^4 + \frac{4}{3}x^3 + \frac{13}{3}x^2 + 4x$ for all $x \ge 0$. Then, define $g(y) : [0, \infty) \mapsto \mathbb{R}$ as:

$$g(y) := y \cdot f\left(2\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + y)\right) \cdot f\left(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + y)\right) \tag{41}$$

So, $g(y)$ is also monotonically increasing on $[0, \infty)$ and $g(0) = 0$.

Lemma 10. *Suppose $Z_{X,0} \in \mathbb{S}^r_{++}$ and $Z_{S,0} \in \mathbb{S}^{n-r}_{--}$. Define two positive constants α_K and C_K:*

$$\alpha_K = \frac{f\left(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + \frac{1}{2})\right)}{\|Z_{X,0}\|_2}, \qquad C_K := \min\{C_1, C_2, C_3, C_4, C_5\}, \tag{42}$$

where

$$C_1 = \frac{1}{2}, \quad C_2 = g^{-1}(\|Z_{X,0}\|_2^2), \quad C_3 = \frac{3}{8\eta_0},$$
$$C_4 = \frac{1}{\sqrt{4\eta_0\alpha_K}}, \quad C_5 = \sqrt{\frac{1}{4\alpha_K}\min\{\lambda_{\min}(Z_{X,0}), -\lambda_{\max}(Z_{S,0})\}}.$$

For any $\|Z_{O,0}\|_2 \le C_K$ and for any integer $\ell \ge 1$, it holds that

$$\|Z_{O,\ell}\|_2 \le \alpha_K \cdot \|Z_{O,0}\|_2^{\ell+1} \tag{43a}$$
$$\|Z_{X,\ell} - Z_{X,0}\|_2 \le \alpha_K \cdot \sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2} \tag{43b}$$
$$\|Z_{S,\ell} - Z_{S,0}\|_2 \le \alpha_K \cdot \sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}. \tag{43c}$$

Moreover, for any integer $\ell \ge 1$, it holds that

$$\frac{2}{3}\eta_0 \le \eta_\ell \le 2\eta_0, \quad \lambda_{\min}(Z_{X,\ell}) \ge \frac{1}{2}\lambda_{\min}(Z_{X,0}) > 0 \quad \lambda_{\max}(Z_{S,\ell}) \le \frac{1}{2}\lambda_{\max}(Z_{S,0}) < 0.$$

Thus, $Z_{X,\ell} \in \mathbb{S}^n_{++}$ and $Z_{S,\ell} \in \mathbb{S}^n_{--}$ for all $\ell \in \mathbb{N}$.

Proof. See Appendix D.2. □

7.2 Step 2: Convergence of $\{V_\ell\}_{\ell=0}^\infty$

Now we show that the sequence $\{V_\ell\}_{\ell=0}^\infty$ converges to $\Pi_{\mathbb{S}^n_+}(Z+H)$; see Lemma 12. This is achieved by bounding the distance between $Q_{\ell+1}$ and Q_ℓ and that between Q_ℓ and $I_n + W_0$; see Lemma 11.

Lemma 11. *For any integer $\ell \ge 1$, it holds that*

$$\|Q_{\ell+1} - Q_\ell\|_2 \le \frac{8}{3}\eta_0\alpha_K \cdot \|Z_{O,0}\|_2^{\ell+1} \tag{44}$$

and

$$\|Q_\ell - (I_n + W_0)\|_2 \le \frac{2}{3}\eta_0^2 \cdot \|Z_{O,0}\|_2^2 + \frac{8}{3}\eta_0\alpha_K \cdot \sum_{i=1}^{\ell-1} \|Z_{O,0}\|_2^{i+1} \tag{45}$$

where η_0 is defined in (40) and α_K in (42).

Proof. See Appendix D.3. □

Lemma 12. *The sequence $\{V_\ell\}_{\ell=1}^\infty$ generated in Algorithm 1 converges to $\Pi_{\mathbb{S}^n_+}(Z+H)$.*

Proof. See Appendix D.4. □

7.3 Step 3: Proof of Theorem 2

Before we execute the last step of our proof, two more details are needed. First, observe that all the three constants, η_0 in (40), α_K and C_K in (42) implicitly rely on $Z_{X,0} = \Lambda_X + H_X$ and $Z_{S,0} = \Lambda_S + H_S$ (though independent of $Z_{O,0} = H_O$). Yet, the constants α_{EB} and C_{EB} in Theorem 2 should be independent of the perturbation H. The uniform bounds of η_0, α_K and C_K is achieved in Lemma 13.

Lemma 13. *Suppose $\|H\|_2 \le \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\}$. Then, it holds that*

$$\lambda_{\min}(Z_{X,0}) \ge \frac{1}{2}\lambda_r > 0 \qquad \lambda_{\max}(Z_{S,0}) \le \frac{1}{2}\lambda_{r+1} < 0.$$

Moreover, there exist three positive constants $\alpha_{K,f}$, $\eta_{0,f}$ and $C_{K,f}$, only depending on n, r and the eigenvalues $\{\lambda_i\}_{i=1}^n$ of Z, such that

$$\eta_0 \le \eta_{0,f}, \qquad \alpha_K \le \alpha_{K,f}, \qquad C_K \ge C_{K,f} > 0.$$

Proof. See Appendix D.5. □

With Lemma 13, as long as $\|H\|_2 \le \min\{C_{K,f}, \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\}\}$, we can safely replace α_K and η_0 in Lemmas 9 to 12 with $\alpha_{K,f}$ and $\eta_{0,f}$.

As the last ingredient, Lemma 14 is needed to control the error in the first Sylvester equation ($\ell = 0$), which only relies on n, r, and the eigenvalues $\{\lambda_i\}_{i=1}^n$ of Z.

Lemma 14. *Suppose that H_X and H_S satisfy $\|H_X\|_2 + \|H_S\|_2 \le \frac{\lambda_r - \lambda_{r+1}}{2nd}$, and that $W_{O,0}$ is the solution for*

$$\begin{aligned} & W_O Z_{X,0} + (-Z_{S,0}) W_O = Z_{O,0} \\ \iff \quad & W_O(\Lambda_X + H_X) - (\Lambda_S + H_S) W_O = H_O. \end{aligned}$$

Then, it holds that

$$\|W_{O,0} - \Theta_0 \circ H_O\|_2 \le \frac{2nd}{(\lambda_r - \lambda_{r+1})^2} \cdot \|H_O\|_2 \cdot (\|H_X\|_2 + \|H_S\|_2),$$

where

$$\Theta_0 = \begin{bmatrix} \frac{1}{\lambda_1 - \lambda_{r+1}} & \cdots & \frac{1}{\lambda_r - \lambda_{r+1}} \\ \vdots & \ddots & \vdots \\ \frac{1}{\lambda_1 - \lambda_n} & \cdots & \frac{1}{\lambda_r - \lambda_n} \end{bmatrix} \in \mathbb{R}^{(n-r)\times r}.$$

Proof. See Appendix D.6. □

To prove Theorem 2, it only remains to upper bound the two terms on the right-hand side of (39) one-by-one. Define C_{EB} as

$$C_{\text{EB}} := \min\left\{C_{K,f},\ \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\},\ \frac{\lambda_r - \lambda_{r+1}}{4nd}\right\}. \tag{46}$$

Note that in the following proof, we have already replaced α_K and η_0 in Lemmas 9 to 12 with $\alpha_{K,f}$ and $\eta_{0,f}$.

1. The first term on the right-hand side of (39) is bounded by

$$\begin{aligned} & \left\| (I_n + W_0) \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} (I_n + W_0)^\mathsf{T} - \Pi_{\mathbb{S}_+^n}(Z) - \Omega \circ H \right\|_2 \\ = & \left\| (I_n + W_0) \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} (I_n + W_0)^\mathsf{T} - \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix} - \Omega \circ H \right\|_2 \\ = & \left\| \begin{bmatrix} Z_{X,\infty} & Z_{X,\infty} W_{O,0}^\mathsf{T} \\ W_{O,0} Z_{X,\infty} & W_{O,0} Z_{X,\infty} W_{O,0}^\mathsf{T} \end{bmatrix} - \begin{bmatrix} \Lambda_X + H_X & \Theta^\mathsf{T} \circ H_O^\mathsf{T} \\ \Theta \circ H_O & 0 \end{bmatrix} \right\|_2 \\ \le & \|Z_{X,\infty} - (\Lambda_X + H_X)\|_2 + \|W_{O,0} Z_{X,\infty} W_{O,0}^\mathsf{T}\|_2 + \|W_{O,0} Z_{X,\infty} - \Theta \circ H_O\|_2 \\ \le & \|Z_{X,\infty} - (\Lambda_X + H_X)\|_2 + \|W_{O,0} Z_{X,\infty} W_{O,0}^\mathsf{T}\|_2 + \|W_{O,0} Z_{X,0} - \Theta \circ H_O\|_2 \\ & + \|W_{O,0}(Z_{X,\infty} - Z_{X,0})\|_2. \end{aligned} \tag{47}$$

Again, we bound the right-hand side of (47) one-by-one.

(a) For the term $\|Z_{X,\infty} - (\Lambda_X + H_X)\|_2$, we have from $Z_{O,0} = H_O$ that

$$
\begin{aligned}
&\|Z_{X,\infty} - (\Lambda_X + H_X)\|_2 = \|Z_{X,\infty} - Z_{X,0}\|_2 \\
=&\left\|\sum_{i=0}^{\infty}(Z_{X,i+1} - Z_{X,0})\right\|_2 \leq \sum_{i=0}^{\infty}\|Z_{X,i+1} - Z_{X,0}\|_2 \\
\leq&\alpha_{K,f} \cdot \sum_{i=0}^{\infty}\|Z_{O,0}\|_2^{i+2} \qquad (48a) \\
=&\alpha_{K,f} \cdot \sum_{i=0}^{\infty}\|H_O\|_2^{i+2} = \alpha_{K,f} \cdot \frac{\|H_O\|_2^2}{1-\|H_O\|_2} \\
\leq&2\alpha_{K,f} \cdot \|H_O\|_2^2, \qquad (48b)
\end{aligned}
$$

where (48a) uses (43a) and (48b) uses $\|H_O\|_2 \leq \|H\|_2 \leq C_{K,f} \leq \frac{1}{2}$.

(b) For the term $\|W_{O,0}Z_{X,\infty}W_{O,0}^{\mathsf{T}}\|_2$, we have

$$\|W_{O,0}Z_{X,\infty}W_{O,0}^{\mathsf{T}}\|_2 \leq \|W_{O,0}\|_2^2 \cdot \|Z_{X,\infty}\|_2 \leq \eta_{0,f}^2 \cdot \|H_O\|_2^2 \cdot \|Z_{X,\infty}\|_2.$$

Since $\|Z_{X,\infty}\|_2 \leq \|Z_\infty\|_2 = \|Z+H\|_2$ is bounded, there exists a positive constant α_1 such that $\|W_{O,0}Z_{X,\infty}W_{O,0}^{\mathsf{T}}\|_2 \leq \alpha_1 \cdot \|H_O\|_2^2$.

(c) For the term $\|W_{O,0}Z_{X,0} - \Theta \circ H_O\|_2$, we have

$$
\begin{aligned}
&\|W_{O,0}Z_{X,0} - \Theta \circ H_O\|_2 \\
=&\|W_{O,0}(\Lambda_X + H_X) - \Theta \circ H_O\|_2 \\
\leq&\|W_{O,0}\Lambda_X - \Theta \circ H_O\|_2 + \|W_{O,0}H_X\|_2 \\
=&\|W_{O,0}\Lambda_X - (H_O \circ \Theta_0)\Lambda_X\|_2 + \|W_{O,0}H_X\|_2 \qquad (49a) \\
\leq&\lambda_1 \cdot \|W_{O,0} - (H_O \circ \Theta_0)\|_2 + \|W_{O,0}H_X\|_2 \\
\leq&\frac{2nd\lambda_1}{(\lambda_r - \lambda_{r+1})^2} \cdot \|H_O\|_2 \cdot (\|H_X\|_2 + \|H_S\|_2) + \|W_{O,0}H_X\|_2 \qquad (49b) \\
\leq&\frac{2nd\lambda_1}{(\lambda_r - \lambda_{r+1})^2} \cdot \|H_O\|_2 \cdot (\|H_X\|_2 + \|H_S\|_2) + \eta_{0,f} \cdot \|H_O\|_2 \cdot \|H_X\|_2 \qquad (49c) \\
\leq&\alpha_2\|H_O\|_2 \cdot (\|H_X\|_2 + \|H_S\|_2) \qquad (49d)
\end{aligned}
$$

for some positive constant α_2. Here, (49a) holds since for a diagonal matrix D:

$$(AD) \circ B = B \circ (AD) = (B \circ A)D,$$

(49b) comes from Lemma 14 since

$$\|H_X\|_2 + \|H_S\|_2 \leq 2\|H\|_2 \leq 2C_{K,f} \leq 2 \cdot \frac{\lambda_r - \lambda_{r+1}}{4nd} = \frac{\lambda_r - \lambda_{r+1}}{2nd},$$

and (49c) follows from Lemma 9 and $\|W_{O,0}\|_2 \leq \eta_{0,f} \cdot \|Z_{O,0}\|_2 = \eta_{0,f} \cdot \|H_O\|_2$.

(d) For the term $\|W_{O,0}(Z_{X,\infty}-Z_{X,0})\|_2$, we have

$$
\begin{aligned}
\|W_{O,0}(Z_{X,\infty}-Z_{X,0})\|_2 &\le \|W_{O,0}\|_2\cdot\|Z_{X,\infty}-Z_{X,0}\|_2\\
&\le \eta_{0,f}\cdot\|H_O\|_2\cdot\|Z_{X,\infty}-Z_{X,0}\|_2\\
&\le \eta_{0,f}\cdot\|H_O\|_2\cdot\alpha_{K,f}\cdot\sum_{i=0}^{\infty}\|H_O\|_2^{i+2} \qquad (50a)\\
&\le \eta_{0,f}\alpha_{K,f}\cdot\|H_O\|_2\cdot\frac{\|H_O\|_2^2}{1-\|H_O\|_2}\\
&\le 2\eta_{0,f}\alpha_{K,f}\cdot\|H_O\|_2^3, \qquad (50b)
\end{aligned}
$$

where (50a) follows from (43b).

2. For the second term on the right-hand side of (39), we see from Lemma 11 that

$$
\begin{aligned}
\|Q_\infty-(I_n+W_0)\|_2 &\le \frac{2}{3}\eta_{0,f}^2\cdot\|H_O\|_2^2+\frac{8}{3}\eta_{0,f}\alpha_{K,f}\cdot\sum_{i=1}^{\infty}\|H_O\|_2^{i+1}\\
&= \frac{2}{3}\eta_{0,f}^2\cdot\|H_O\|_2^2+\frac{8}{3}\eta_{0,f}\alpha_{K,f}\cdot\frac{\|H_O\|_2^2}{1-\|H_O\|_2}\\
&\le \frac{2}{3}\eta_{0,f}^2\cdot\|H_O\|_2^2+\frac{16}{3}\eta_{0,f}\alpha_{K,f}\cdot\|H_O\|_2^2.
\end{aligned}
$$

Together with the boundedness of $\|I_n+W_0\|_2$ and $Z_{X,\infty}$, we conclude that there exists a positive constant α_3 such that

$$
\begin{aligned}
&\left\|Q_\infty\begin{bmatrix}Z_{X,\infty}&0\\0&0\end{bmatrix}Q_\infty^{\mathsf{T}}-(I_n+W_0)\begin{bmatrix}Z_{X,\infty}&0\\0&0\end{bmatrix}(I_n+W_0)^{\mathsf{T}}\right\|_2\\
&\le 2\left\|(Q_\infty-(I_n+W_0))\begin{bmatrix}Z_{X,\infty}&0\\0&0\end{bmatrix}(I_n+W_0)^{\mathsf{T}}\right\|_2\\
&\quad+\left\|(Q_\infty-(I_n+W_0))\begin{bmatrix}Z_{X,\infty}&0\\0&0\end{bmatrix}(Q_\infty-(I_n+W_0))^{\mathsf{T}}\right\|_2\\
&\le \alpha_3\cdot\|H_O\|_2^2. \qquad (51)
\end{aligned}
$$

Therefore, combining (39), (47), (48b), (49d), (50b) and (51) yields

$$
\begin{aligned}
&\|\Pi_{\mathbb{S}_+^n}(Z+H)-\Pi_{\mathbb{S}_+^n}(Z)-\Omega\circ H\|_2\\
&\le 2\alpha_{K,f}\cdot\|H_O\|_2^2+\alpha_2\cdot\|H_O\|_2^2+\alpha_2\cdot\|H_O\|_2\cdot(\|H_X\|_2+\|H_S\|_2)\\
&\quad+2\eta_{0,f}\alpha_{K,f}\cdot\|H_O\|_2^2+\alpha_3\cdot\|H_O\|_2^2\\
&\le \alpha_{\mathrm{EB}}\cdot\|H_O\|_2\cdot\|H\|_2
\end{aligned}
$$

for some positive constant α_{EB}. This concludes the proof.

7.4 Generalization to the Non-diagonal Case

Though the previous analysis is performed under the assumption that Z is a diagonal matrix, straightforward computation generalizes our result to the more general, non-diagonal case. When $Q \neq I_n$, we have

$$
\begin{aligned}
& \|\Pi_{\mathbb{S}_+^n}(Z+H) - \Pi_{\mathbb{S}_+^n}(Z) - Q(Q^\top H Q)Q^\top\|_2 \\
= & \|Q(Q^\top \Pi_{\mathbb{S}_+^n}(Z+H)Q - Q^\top \Pi_{\mathbb{S}_+^n}(Z)Q - Q^\top H Q)Q^\top\|_2 \\
= & \|Q^\top \Pi_{\mathbb{S}_+^n}(Z+H)Q - Q^\top \Pi_{\mathbb{S}_+^n}(Z)Q - Q^\top H Q\|_2 && \text{(52a)} \\
= & \|\Pi_{\mathbb{S}_+^n}(Q^\top Z Q + Q^\top H Q) - \Pi_{\mathbb{S}_+^n}(Q^\top Z Q) - Q^\top H Q\|_2 && \text{(52b)} \\
= & \|\Pi_{\mathbb{S}_+^n}(Q^\top Z Q + \widetilde{H}) - \Pi_{\mathbb{S}_+^n}(Q^\top Z Q) - \widetilde{H}\|_2 \\
\leq & \alpha_{\mathrm{EB}} \cdot \|\widetilde{H}_O\|_2 \cdot \|\widetilde{H}\|_2, && \text{(52c)}
\end{aligned}
$$

where (52a) follows from the fact that $\|QA\|_2 = \|A\|_2$, for any matrix A, (52b) uses $\Pi_{\mathbb{S}_+^n}(Q^\top X Q) = Q^\top \Pi_{\mathbb{S}_+^n}(X)Q$, and (52c) holds since $Q^\top Z Q$ is diagonal.

8 Numerical Experiments

In this section, numerical evidence is reported to support our theoretical findings. In particular, numerical experiments are conducted to demonstrate the following.

1. Local (R-)linear convergence is observed, regardless of the (non)degeneracy of the SDP.
2. The established (R-)linear rate of convergence (*e.g.,* in Theorem 3 and Lemma 5) is numerically tight.
3. When SC is close to failure, ADMM for SDP may be extremely slow and no clear linear convergence can be observed within the stated computational budget.

Experiments are performed on a high-performance workstation equipped with a 2.7 GHz AMD 64-Core sWRX8 Processor and 1 TB of RAM. For the standard SDP (1), we denote primal infeasibility r_{p}, dual infeasibility r_{d}, and relative gap r_{gap} as:

$$
r_{\mathrm{p}} := \frac{\|\mathcal{A}X - b\|_2}{1 + \|b\|_2}, \quad r_{\mathrm{d}} := \frac{\|\mathcal{A}^* y + S - C\|_{\mathsf{F}}}{1 + \|C\|_{\mathsf{F}}}, \quad r_{\mathrm{gap}} := \frac{\left|\langle C, X\rangle - b^\top y\right|}{1 + |\langle C, X\rangle| + |b^\top y|},
$$

and define the maximum KKT residual $r_{\max} := \max\{r_{\mathrm{p}}, r_{\mathrm{d}}, r_{\mathrm{gap}}\}$. Unless specified, the stopping criteria are $r_{\max} \leq 10^{-10}$, or the maximum iteration number goes beyond 10^6, or the CPU time exceeds 100 hours. Table 1 presents the data for all the tested SDP instances. The strict complementarity condition is checked by computing $\lambda_{\min}(|Z_\star|)$, the smallest eigenvalues of $Z_\star$ in absolute values.

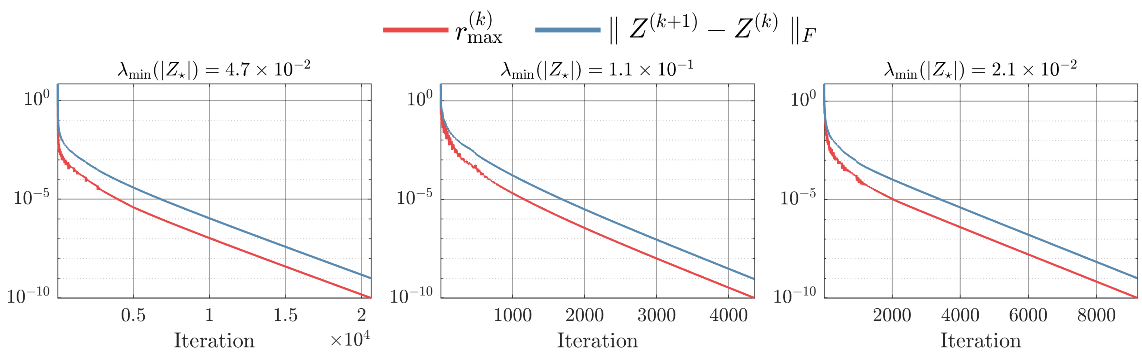
8.1 Demonstration of Local Linear Convergence

In this section, we solve a considerable number of SDPs arising from various applications. In all the experiments, local linear convergence of ADMM is clearly observed, regardless of the (non)degeneracy of the SDPs.

- **MAXCUT [14].** Figure 2 reports three representative examples. In all three cases, strict complementarity holds numerically and ADMM enters the linear convergence region rather quickly.
- **Hamming set problems [48].** Figure 3 reports three representative examples. In all three cases, strict complementarity holds numerically.
- **Maximum stable set problems [42].** Figure 4 reports three representative examples. In all three cases, strict complementarity holds numerically.

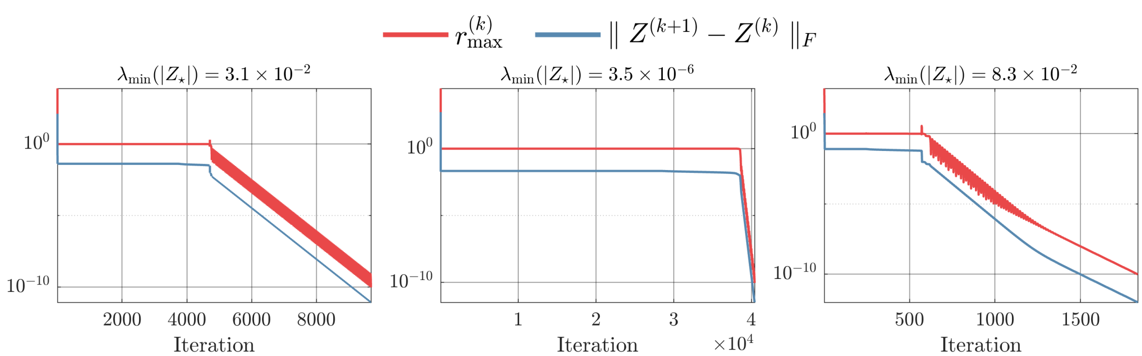
	n	m	σ		n	m	σ
MAXCUT-G*	800	800	1	hamming-9-8	512	2305	0.01
hamming-10-2	1024	23041	0.01	hamming-11-2	2048	56321	0.01
hamming-7-5-6	128	1793	0.01	hamming-8-3-4	256	16129	0.01
hamming-9-5-6	512	53761	0.01	theta-12	600	17979	1
theta-102	500	37467	1	theta-123	600	90020	1
XM-48	144	241	100	XM-93	279	466	100
XM-149	447	746	100	BQP-r*-20-*	231	20601	100
BQP-r*-30-*	496	91326	100	BQP-r*-40-*	861	296001	100
QS-20	231	16402	100	QS-30	496	77377	100
QS-40	861	236202	100	Quasar-100	404	31301	100
Quasar-200	804	122601	100	Quasar-500	2004	756501	100
swissroll	800	3380	1	cnhil10	220	5005	0.01
1dc-1024	1024	24064	100	rose13	105	2379	1
neosfbr25	577	14376	1				

Table 1: Details about all the tested SDP instances. n is the size of the matrix, m is the number of equality constraints, and σ is the fixed penalty paramater in ADMM.



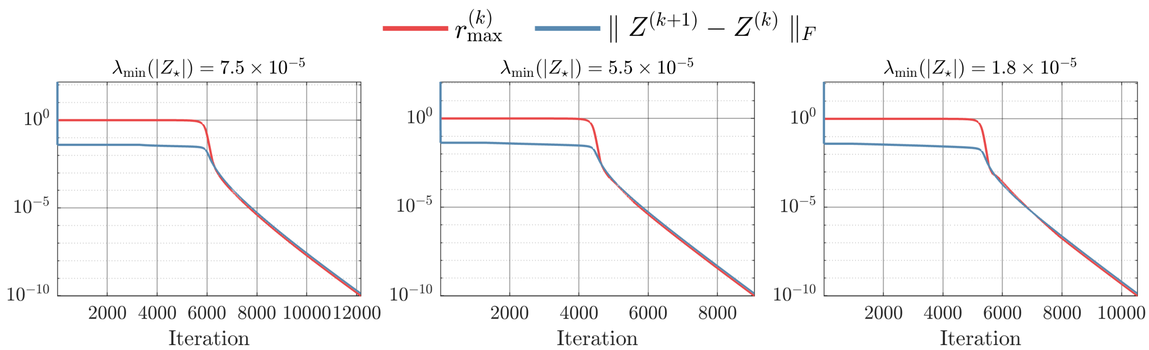
MAXCUT-G1 MAXCUT-G9 MAXCUT-G18

Figure 2: MAXCUT problems with with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.



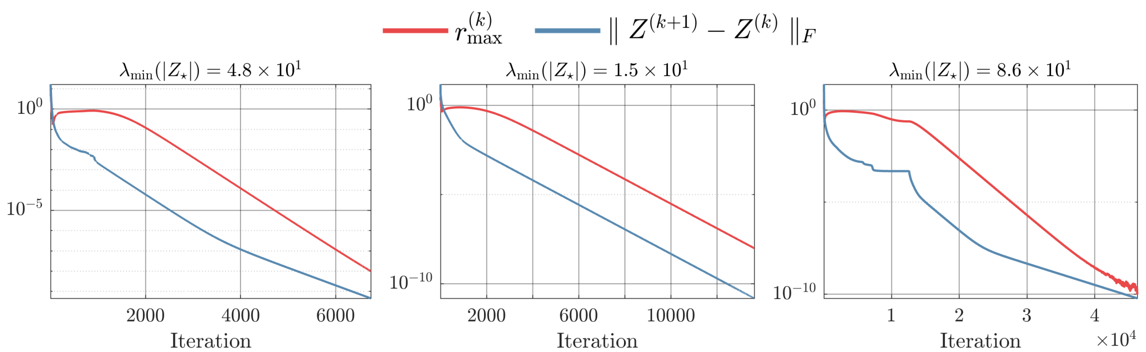
hamming-9-8 hamming-11-2 hamming-7-5-6

Figure 3: Additional Hamming graph problems with with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.



theta-12 theta-102 theta-123

Figure 4: Maximum stable set problems with with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.



XM-48 XM-93 XM-149

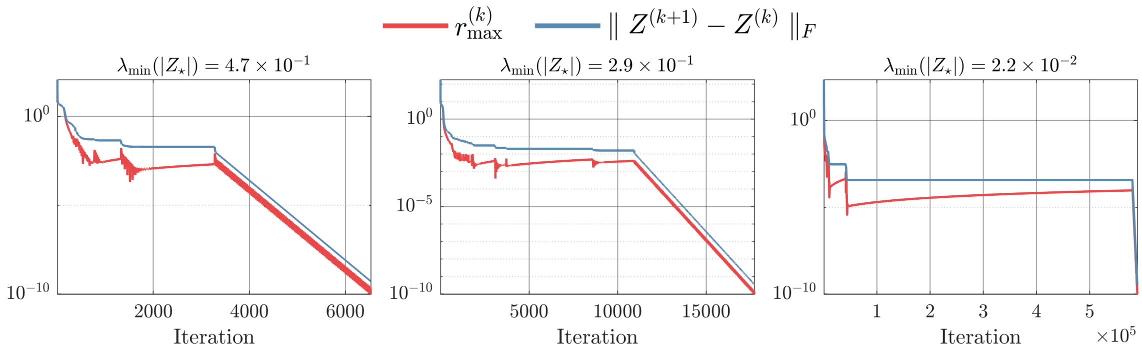
Figure 5: Structure-from-motion problems with with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.

- **Structure from motion problems [23].** Figure 5 reports three representative examples. In all three cases, strict complementarity holds numerically. For XM-48 and XM-93, the maximum KKT residual $r_{\max}$ can only reach 10^{-8} due to numerical errors from eigenvalue decomposition and the unbalance between infeasibility and relative gap.
- **Binary quadratic programming (BQP).** We consider the second-order moment-sum-of-squares (moment-SOS) relaxation [32] of the following polynomial optimization problem

$$\begin{array}{ll} \text{minimize} & \frac{1}{2}x^\mathsf{T}Qx + c^\mathsf{T}x \\ \text{subject to} & 1 - x_i^2 = 0, \quad i \in [n], \end{array}$$

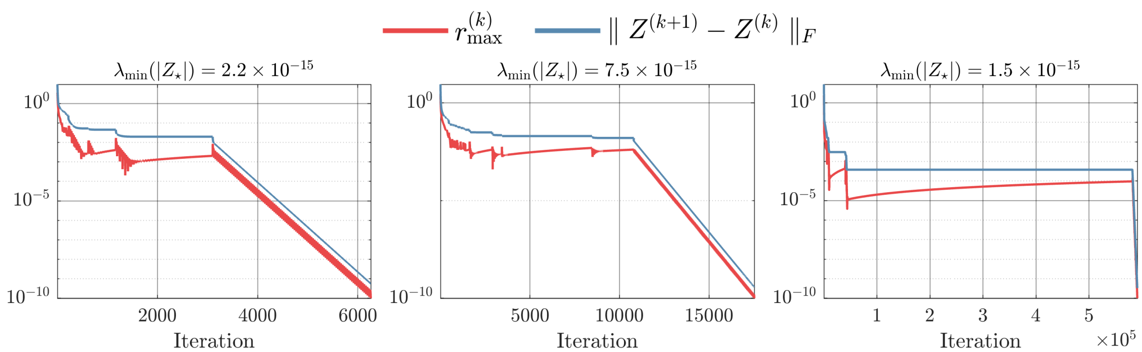
  where the optimization variable is $x \in \mathbb{R}^n$, and the data are $Q \in \mathbb{S}^n$ and $c \in \mathbb{R}^n$. Depending on the data, nondegeneracy (ND) and strict complementarity (SC) conditions may or may not hold.

  - **Case 1: primal ND fails and SC holds.** When $c \sim \mathcal{N}(0, I_n)$, the SDP relaxation is empirically tight and the primal optimal solution has rank one [56, 63]. In this case, primal nondegeneracy fails. Figure 6 reports three representative examples with random (standard Gaussian) initial guess.
  - **Case 2: primal ND fails and SC fails.** We test the same BQP instances with all-zeros initialization. As shown in Figure 7, strict complementarity seems to fail. Nonetheless, the failure of strict complementarity does not deteriorate the linear convergence rate.
  - **Case 3: both primal and dual ND fail and SC holds.** When $c = 0$, the SDP relaxation is



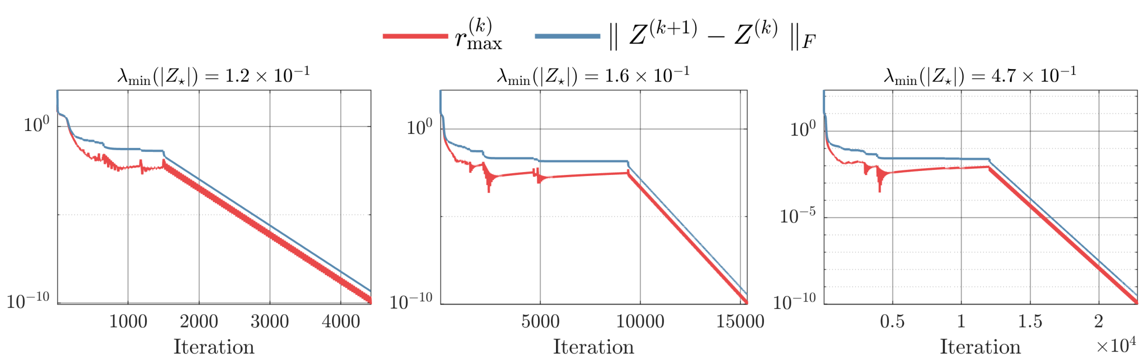
BQP-r1-20-1 BQP-r1-30-1 BQP-r1-40-1

Figure 6: Random BQP problems with $c \sim \mathcal{N}(0, I_n)$ with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.



BQP-r1-20-1 BQP-r1-30-1 BQP-r1-40-1

Figure 7: Random BQP problems with $c \sim \mathcal{N}(0, I_n)$ with all-zero initial guess. In all cases, the converging $Z_\star$ is singular.



BQP-r2-20-1 BQP-r2-30-1 BQP-r2-40-1

Figure 8: random BQP problems with $c = 0$ with random (standard Gaussian) initial guess, under which both primal and dual nondegeneracy fail. In all cases, the converging $Z_\star$ is nonsingular.

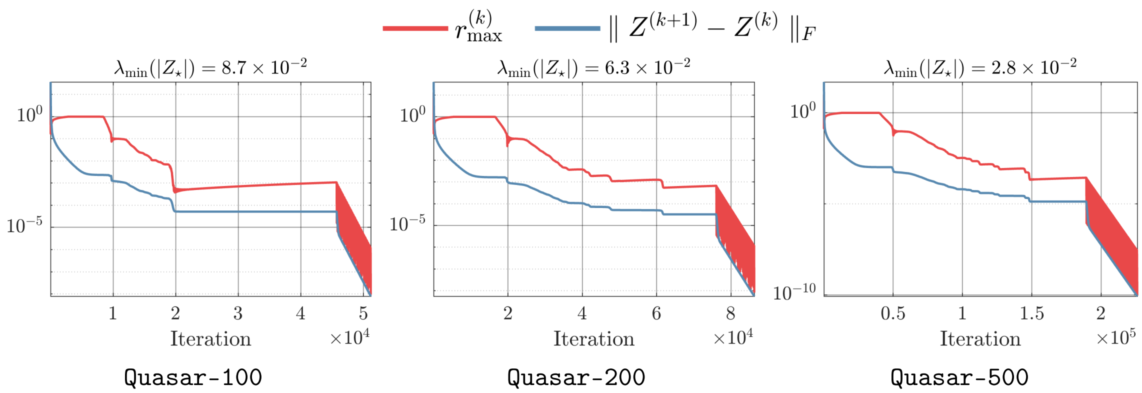


Figure 9: Quasar problems with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.

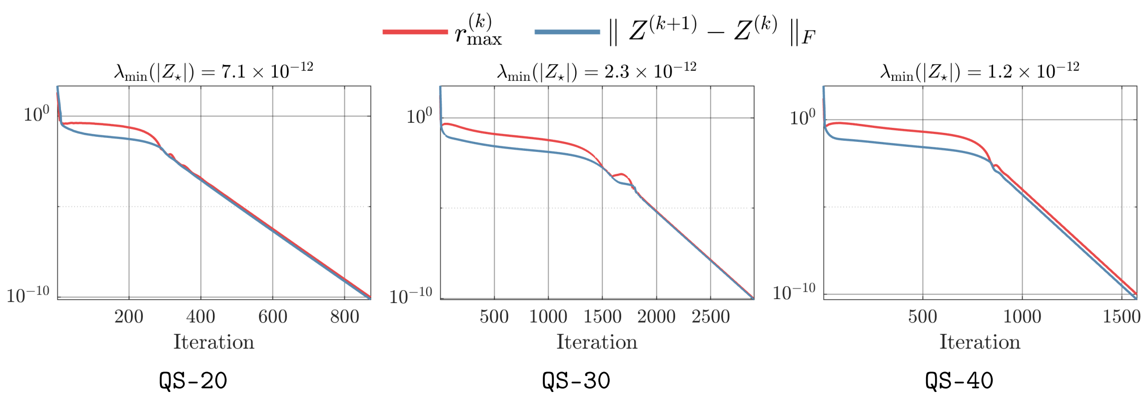


Figure 10: Random QS problems with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is singular.

still empirically tight [56, 63]. However, the primal optimal solution is no longer unique (due to sign symmetry). In this case, both primal and dual nondegeneracy fail and linear convergence is still observed; see Figure 8.

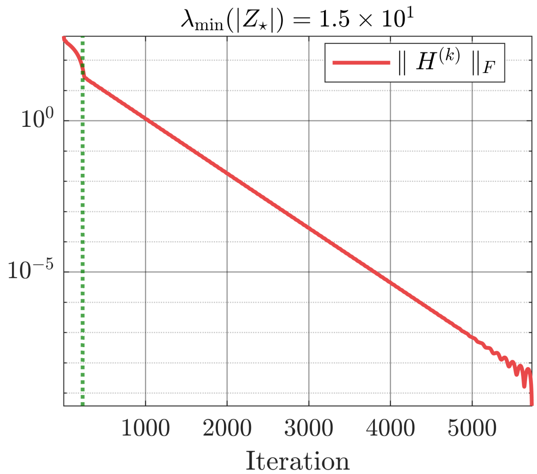
- **Quasar problems [61].** In Quasar problems, the primal solution is unique and has rank one. Similar to BQP, primal nondegeneracy always fails in Quasar problems [61]. Figure 9 reports three examples, in which strict complementarity holds numerically.
- **Quartic function over sphere (QS).** Another classical polynomial optimization problem [56, 63]. In its second-order relaxation, the primal solution is unique and has rank one. Similar to BQP, primal nondegeneracy of QS always fails. In comparison, Figure 10 reports three representative examples. In these cases, strict complementarity seems to fail numerically, but linear convergence is still observed.

Additional numerical results can be found in Appendix E.

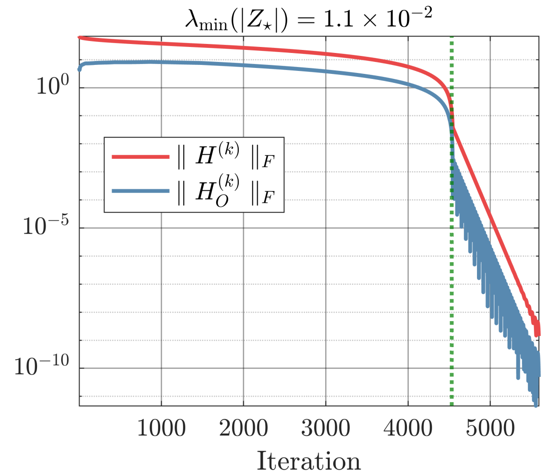
8.2 Demonstration of Numerical Rates

In this section, we numerically verify that the tightness of the derived (R-)linear rate of convergence. In the following two experiments, primal and dual nondegeneracy are checked numerically as follows. We compute

$$
\begin{aligned}
W_1 &:= \begin{bmatrix} \operatorname{svec}(A_1) & \operatorname{svec}(A_2) & \cdots & \operatorname{svec}(A_m) \end{bmatrix} \\
W_2 &:= \begin{bmatrix} \cdots & \operatorname{svec}\left(Q_\star E_{i,j} Q_\star^\top\right) & \cdots \end{bmatrix}, \quad \text{for} i = r+1, \dots, n, \text{ and } j = i, \dots, n,
\end{aligned}
$$



(a) ND holds and SC holds



(b) ND fails and SC holds

Figure 11: Demonstration of numerical rates. (a) Plot of $\|H_O^{(k)}\|_\mathsf{F}$ from a toy structure-from-motion problem. (b) Plot of both $\|H^{(k)}\|_\mathsf{F}$ and $\|H_O^{(k)}\|_\mathsf{F}$ from a toy BQP problem. In both cases, the numerical rates match quite well with the theory; see Theorem 3 and Lemma 5.

where $E_{i,j} \in \mathbb{R}^{n\times n}$ is the (i,j)th elementary matrix; *i.e.*, all the elements are zero except the (i,j)th entry is one. It is clear that the columns of $W_1 \in \mathbb{R}^{t(n)\times m}$ form a basis of $\mathcal{R}(\mathcal{A}^*)$ and those of $W_2 \in \mathbb{R}^{t(n)\times t(n-r)}$ form a basis of $\mathcal{T}_{X_\star}^\perp$. To check primal nondegeneracy $\mathcal{R}(\mathcal{A}^*) \cap \mathcal{T}_{X_\star}^\perp = \{0\}$, it suffices to check the following rank condition:

$$\operatorname{rank}(W_1) + \operatorname{rank}(W_2) = \operatorname{rank}(\begin{bmatrix} W_1 & W_2 \end{bmatrix}). \tag{53}$$

Dual nondegeneracy can be checked in a similar manner.

(a) **ND fails and SC holds.** In Figure 11 (a), we consider a toy problem from structure-from-motion dataset with 15 frames. In this case, the matrix size is $n = 15$ and the numerical ranks are

$$\operatorname{rank}(W_1) = 903, \qquad \operatorname{rank}(W_2) = 76, \qquad \operatorname{rank}(\begin{bmatrix} W_1 & W_2 \end{bmatrix}) = 979,$$

which satisfies the condition (53), and thus primal nondegeneracy holds. Dual nondegeneracy is verified similarly. In this nondegenerate case, we see from Theorem 3 that the sequence $\|H^{(k)}\|_\mathsf{F}$ converges linearly with rate $\|\mathcal{M}\|_{\mathrm{op}} = 0.998$, which matches quite well with the numerical rate 0.996 from Figure 11 (a).

(b) **ND fails and SC holds.** In Figure 11 (b), we consider a toy BQP problem with 10 binary variables. So, the matrix size is $n = 66$ and the numerical ranks are

$$\operatorname{rank}(W_1) = 1826, \qquad \operatorname{rank}(W_2) = 2145, \qquad \operatorname{rank}(\begin{bmatrix} W_1 & W_2 \end{bmatrix}) = 2211,$$

which implies the failure of primal nondegeneracy. Similarly, dual nondegeneracy holds numerically. From Lemma 5, the sequence $\|H_O^{(k)}\|_\mathsf{F}$ converges R-linearly with rate $\|\mathcal{M} - \Pi_{\mathrm{Fix}(\mathcal{M})}\|_{\mathrm{op}} = 0.984$. As expected, the numerical rate from Figure 11 is also 0.984, which suggests the tightness of our theory. More interestingly, the numerical rate of $\|H_O^{(k)}\|_\mathsf{F}$ and that of $\|H^{(k)}\|_\mathsf{F}$ are exactly the same in this example.

8.3 Failure Cases

In Figure 12, we report some SDP instances for which ADMM fails to achieve $r_{\max} \leq 10^{-10}$ within the stated budget and no clear linear convergence is observed. A common feature in these instances is that the values $\lambda_{\min}(|Z_\star|)$ tend to be small ($e.g.$, $10^{-4} \sim 10^{-9}$), yet not exactly zero (compared to QS and BQP cases where $\lambda_{\min}(|Z_\star|) < 10^{-14}$). The near failure of strict complementarity may qualitatively explain the lack of linear convergence. Recall from Theorem 2 that the refined error bound holds for "small" perturbation $\|H\|_2 \leq C_{\mathrm{EB}}$, which is proportional to $\min\{\lambda_r, -\lambda_{r+1}\}$; see (46). Consequently, a small $\lambda_{\min}(|Z_\star|) = \min\{\lambda_r, -\lambda_{r+1}\}$ enforces a tiny perturbation radius C_{EB}. So, linear convergence of ADMM, if exists, must occur at a very late stage. This observation also aligns with the recent findings in first-order methods for LP [40].

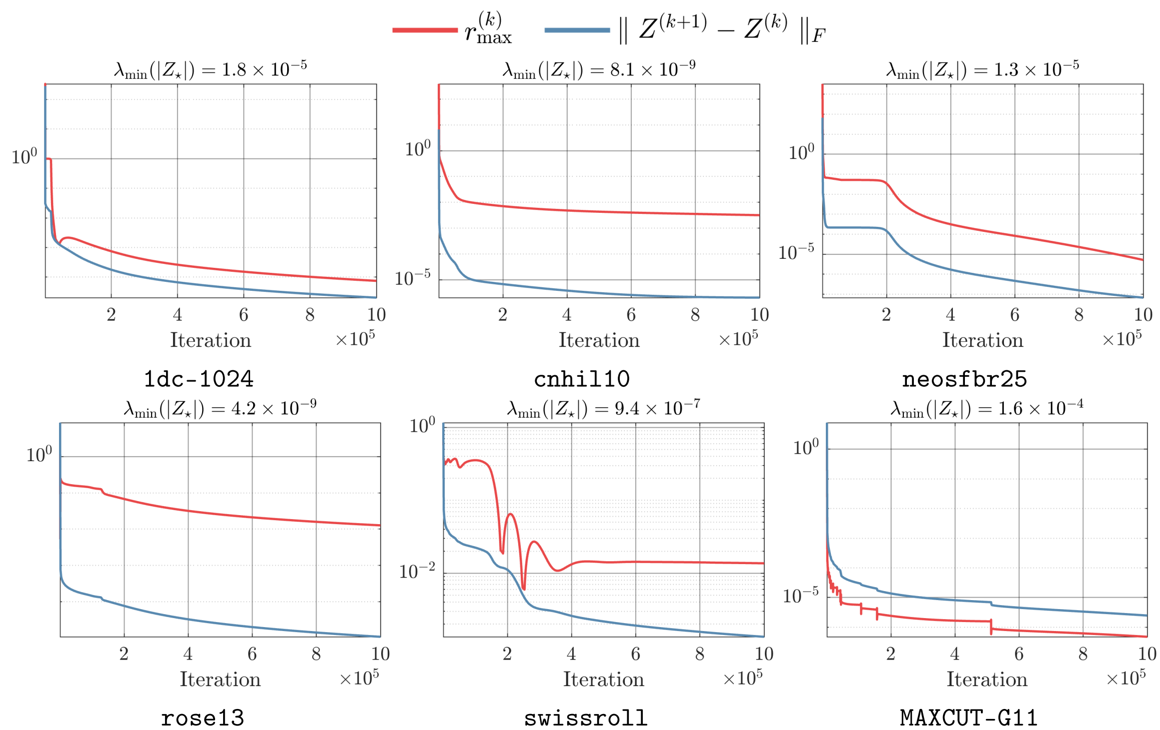


Figure 12: Cases from datasets [14, 42]. For these SDPs, ADMM fails to achieve $r_{\max} \leq 10^{-10}$ within the stated budget, and no clear linear convergence is observed.

9 Discussion: Rank Identification and Linear Convergence

First-order methods ($e.g.,$ PDHG) for LP (a special case of SDP) are known to have an intriguing two-stage phenomenon [40]. The first stage identifies the basis and finishes in a finite number of iterations, with a sublinear rate. Then, the second stage of the algorithm converges linearly, with a rate related to the local sharpness constant. In view of the equivalence between ADMM and PDHG [45], as well as the similarity between the numerical results in Section 8 and those in [40], it is natural to ask whether ADMM for SDP has a similar two-stage performance and whether it could identify the solution *rank* (*c.f.*, basis in LP) within a finite number of iterations. This section aims to provide a partial answer to the above questions, both theoretically and empirically.

Finite-time rank identification. In the context of ADMM for SDP, the fact that rank identification occurs within a finite number of iterations is readily guaranteed by the well-known partial smoothness theory [18, 33, 60]. More precisely, rank identification means that ADMM identifies the rank of the solution it converges to and all the subsequent ADMM iterates have the same rank. Here, we provide a more direct proof in the context of SDP, without invoking the more general partial smoothness theory.

Proposition 3. *Suppose that Assumptions 1 and 2 hold and that ADMM (2) converges to $(X_\star, y_\star, S_\star)$. Then, there exists $\bar{k}_{\mathrm{ID}} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}_{\mathrm{ID}}$, it holds that*

$$\operatorname{rank}(X^{(k)}) = \operatorname{rank}(X_\star), \qquad \operatorname{rank}(S^{(k)}) = \operatorname{rank}(S_\star).$$

Proof. First, we show that

$$\operatorname{rank}(\Pi_{\mathbb{S}^n_+}(Z_\star + H^{(k)})) = \operatorname{rank}(\Pi_{\mathbb{S}^n_+}(Z_\star)) \quad \text{if } \|H^{(k)}\|_2 < \min\{\lambda_r, -\lambda_{r+1}\}.$$

To see this, denote by γ_r and γ_{r+1} the rth and $(r+1)$st largest eigenvalue of $Z_\star + H^{(k)}$, respectively. Then, by Weyl's inequality, we have

$$\gamma_r \geq \lambda_r - \|H^{(k)}\|_2 > \lambda_r - \min\{\lambda_r, -\lambda_{r+1}\} \geq 0,$$
$$\gamma_{r+1} \leq \lambda_{r+1} + \|H^{(k)}\|_2 < \lambda_{r+1} + \min\{\lambda_r, -\lambda_{r+1}\} \leq 0,$$

where recall λ_r and λ_{r+1} are the rth and $(r+1)$st largest eigenvalue of $Z_\star$, respectively. Thus, we have $\gamma_r > 0 > \gamma_{r+1}$ and

$$\operatorname{rank}(X^{(k)}) = \operatorname{rank}(\Pi_{\mathbb{S}^n_+}(Z_\star + H^{(k)})) = r = \operatorname{rank}(\Pi_{\mathbb{S}^n_+}(Z_\star)).$$

The dual part follows in a symmetric manner:

$$\operatorname{rank}(S^{(k)}) = \operatorname{rank}(\Pi_{\mathbb{S}^n_+}(-Z_\star - H^{(k)})) = n - r = \operatorname{rank}(\Pi_{\mathbb{S}^n_+}(-Z_\star)).$$

Second, since $\|H^{(k)}\|_2 \to 0$ as $k \to \infty$, there exists $\bar{k}_{\mathrm{ID}} \in \mathbb{N}$ such that $\|H^{(k)}\|_2 < \min\{\lambda_r, \lambda_{r+1}\}$. This concludes the proof. ☐

On the relation between rank identification and linear convergence. Considering both rank identification and local linear convergence, it is natural to investigate the relationship of these two phenomena: which one occurs first? Unlike the case of PDHG for LP, it remains unclear whether rank identification is the trigger for linear convergence.

Here, we provide a simple example that to some extent explains the interaction between these two phenomena. Recall from our analysis (specifically Lemma 6) that the R-linear convergence of $\|H^{(k)}\|_{\mathsf{F}}$ is built upon that of the two sequences

$$\|\Pi_{\mathcal{T}_{S_\star}}(X^{(k)})\|_{\mathsf{F}} = \mathcal{O}(\|H_O^{(k)}\|_{\mathsf{F}}), \qquad \|\Pi_{\mathcal{T}_{X_\star}}(S^{(k)})\|_{\mathsf{F}} = \mathcal{O}(\|H_O^{(k)}\|_{\mathsf{F}}). \tag{54}$$

In view of this, we build an SDP instance in which rank identification does not occur and (54) fails to hold. So, in the worst case, (54) needs rank identification. If (54) were necessary for linear convergence, then we could conclude that rank identification occurs no later than the final (R-)linear convergence regime.

Example 1. *Consider the SDP (1) with $n = 3$. Suppose Assumption 1, primal nondegeneracy (8) and dual nondegeneracy (9) hold. Suppose* $\operatorname{rank}(X_\star) = 1$ *and* $\operatorname{rank}(S_\star) = 2$. *(So Assumption 2 holds.) Suppose $Z_\star =$* $\operatorname{diag}(1, -\delta, -\delta)$, *where $\delta > 0$ can be arbitrarily small. Then, Proposition 3 implies that rank identification must occur if $\|H^{(k)}\|_2 < \delta$.*

Assume, without loss of generality, that ADMM starts at the following points (with $\epsilon > 0$)

$$X^{(0)} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{\epsilon}{2} & \frac{\epsilon}{2} \\ 0 & \frac{\epsilon}{2} & \frac{\epsilon}{2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\frac{\epsilon}{2}} \\ 0 & \sqrt{\frac{\epsilon}{2}} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{\frac{\epsilon}{2}} \\ 0 & \sqrt{\frac{\epsilon}{2}} \end{bmatrix}^{\mathsf{T}},$$

$$\sigma S^{(0)} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \delta + \frac{\epsilon}{2} & -\delta - \frac{\epsilon}{2} \\ 0 & -\delta - \frac{\epsilon}{2} & \delta + \frac{\epsilon}{2} \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{\delta + \frac{\epsilon}{2}} \\ -\sqrt{\delta + \frac{\epsilon}{2}} \end{bmatrix} \begin{bmatrix} 0 \\ \sqrt{\delta + \frac{\epsilon}{2}} \\ -\sqrt{\delta + \frac{\epsilon}{2}} \end{bmatrix}^{\mathsf{T}}.$$

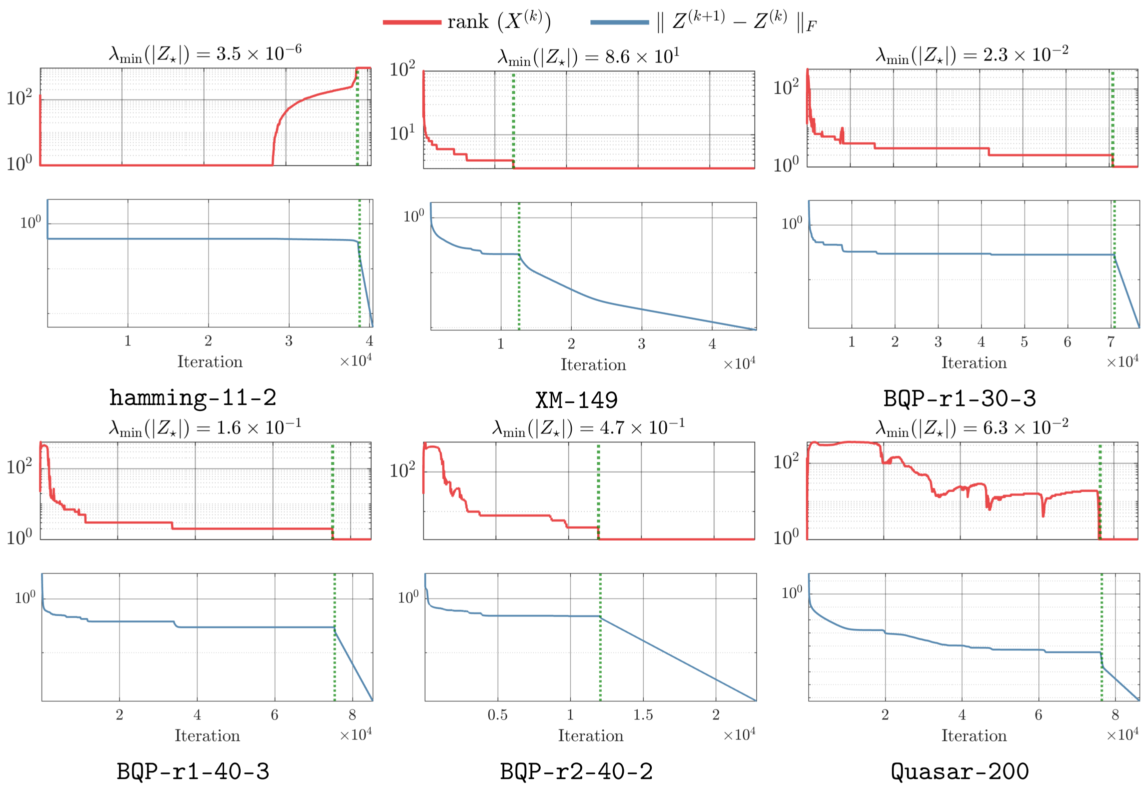


Figure 13: Six representative SDP instances illustrating rank identification: almost at the same time when $X^{(k)}$ identifies the solution rank, ADMM steps into the final linear convergence region.

It is clear that $\operatorname{rank}(X^{(0)}) = 1$, $\operatorname{rank}(S^{(0)}) = 2$, $\langle X^{(0)}, S^{(0)} \rangle = 0$, *and*

$$H^{(0)} = X^{(0)} - \sigma S^{(0)} - Z_\star = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \delta + \epsilon \\ 0 & \delta + \epsilon & 0 \end{bmatrix}.$$

Moreover, $\|H^{(0)}\|_2 = \delta + \epsilon$ *and* $H_O^{(0)} = 0$*. On the other hand,*

$$\Pi_{\mathcal{T}_{S_\star}}(X^{(0)}) = \begin{bmatrix} 0 & 0 & 0 \\ 0 & \frac{\epsilon}{2} & \frac{\epsilon}{2} \\ 0 & \frac{\epsilon}{2} & \frac{\epsilon}{2} \end{bmatrix}.$$

To conclude, rank identification does not occur and (54) *fails to hold.*

Numerical evidence. As shown in Figure 13, for many tested SDP instances, as soon as $X^{(k)}$ identifies the solution rank, the ADMM iterates simultaneously steps into the final region of linear convergence.

In view of Proposition 3, Example 1 and Figure 13, one may already identify a gap between theory and practice.

> **Open problems**: In what type of SDPs is rank identification a necessary condition for (R-)linear convergence? Under which conditions will rank identification and (R-)linear convergence occur simultaneously?

10 Conclusion

We established a new sufficient condition for the local linear convergence of the Alternating Direction Method of Multipliers (ADMM) in solving semidefinite programming (SDP) problems. Contrary to the conventional belief that ADMM is inherently slow for SDPs, we demonstrated that when the converged primal–dual optimal solutions satisfy strict complementarity, ADMM exhibits local linear convergence, regardless of nondegeneracy conditions. Our theoretical analysis is grounded in a direct local linearization of the ADMM operator and a refined error bound for the projection onto the positive semidefinite cone, revealing the anisotropic nature of projection residuals and improving previous bounds.

Extensive numerical experiments validated our theoretical findings, showing that ADMM achieves local linear convergence across a variety of SDP instances, including those where nondegeneracy fails. Furthermore, we identified cases where ADMM struggles to reach high accuracy, linking these difficulties to near violations of strict complementarity. This observation aligns with recent results in linear programming.

Our numerical results also revealed intriguing connections between rank identification and linear convergence. While we provided a qualitative analysis, a complete understanding remains open. Future work could further investigate this relationship, examine whether linear convergence can occur in the absence of both nondegeneracy and strict complementarity, and develop new algorithms to accelerate ADMM and other first-order methods.

Acknowledgements

We sincerely thank Ying Cui, Jingwei Liang, Ling Liang, Feng-Yi Liao, Haihao Lu, Defeng Sun, and Tianyun Tang for their valuable discussions. Xin Jiang extends special gratitude to Adrian S. Lewis for his insightful perspectives and unwavering support.

Appendix A Discussion on [12, Proposition 3.4]

In this section, we show that [12, Proposition 3.4], under an additional nonsingularity assumption, can be readily derived from Theorem 2. We first restate [12, Proposition 3.4] (with the nonsingularity assumption) below.

Corollary 1. *Let $Z \in \mathbb{S}^n$ being nonsingular and $X, S \in \mathbb{S}^n_+$ be defined as*

$$Z := \begin{bmatrix} \Lambda_X & 0 \\ 0 & \Lambda_S \end{bmatrix}, \qquad X := \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix}, \qquad S := \begin{bmatrix} 0 & 0 \\ 0 & -\Lambda_S \end{bmatrix},$$

where, without loss of generality, $\Lambda_X \in \mathbb{R}^{r \times r}_{++}$ and $\Lambda_S \in \mathbb{R}^{s \times s}_{++}$ are diagonal matrices of the form in (6), and $r + s = n$. For a sufficiently small perturbation $H \in \mathbb{S}^n$, define

$$X' := \Pi_{\mathbb{S}^n_+}(Z + H), \qquad S' := X' - (Z + H), \qquad \Delta X := X' - X, \qquad \Delta S := S' - S.$$

Denote $\alpha := \{1, 2, \ldots, r\}$ and $\gamma := \{r + 1, \ldots, n\}$. Then, it holds that

$$\begin{aligned} X'_{\alpha\alpha} &= \Lambda_X + \mathcal{O}(\|\Delta X\|), & S'_{\gamma\gamma} &= \Lambda_S + \mathcal{O}(\|\Delta S\|), \\ X'_{\gamma\alpha} &= \mathcal{O}(\min\{\|\Delta X\|, \|\Delta S\|\}), & S'_{\gamma\alpha} &= \mathcal{O}(\min\{\|\Delta X\|, \|\Delta S\|\}), \\ X'_{\gamma\gamma} &= \mathcal{O}(\|\Delta X\| \cdot \|\Delta S\|), & S'_{\alpha\alpha} &= \mathcal{O}(\|\Delta X\| \cdot \|\Delta S\|), \end{aligned}$$
$$S'_{\gamma\alpha}\Lambda_X - \Lambda_S X'_{\gamma\alpha} = \mathcal{O}(\|\Delta X\| \cdot \|\Delta S\|),$$

where $\|\cdot\|$ is an arbitrary matrix norm.

Proof. When $\|H\|$ is sufficiently small, we conclude from Theorem 2 that there exists two positive constants κ_X and κ_S (depending on the norm type) such that

$$\|\Delta X - \Omega \circ H\| = \left\| \begin{bmatrix} X'_{\alpha\alpha} - \Lambda_X - H_X & X'_{\alpha\gamma} - \Theta^\mathsf{T} \circ H_O^\mathsf{T} \\ X'_{\gamma\alpha} - \Theta \circ H_O & X'_{\gamma\gamma} \end{bmatrix} \right\| \le \kappa_X \cdot \|H_O\| \cdot \|H\|,$$
$$\begin{aligned} \|\Delta S - (E_n - \Omega) \circ H\| &= \left\| \begin{bmatrix} S'_{\alpha\alpha} & S'_{\alpha\gamma} - (E_{(n-r)\times r} - \Theta)^\mathsf{T} \circ H_O^\mathsf{T} \\ S'_{\gamma\alpha} - (E_{(n-r)\times r} - \Theta) \circ H_O & S'_{\gamma\gamma} - \Lambda_S - H_S \end{bmatrix} \right\| \\ &\le \kappa_S \cdot \|H_O\| \cdot \|H\|. \end{aligned}$$

Thus, there exist four positive constraints $\kappa_1, \kappa_2, \kappa_3, \kappa_4$ such that

$$\kappa_1 \cdot (\|H_X\| + \|H_O\|) \le \|\Delta X\| \le \kappa_2 \cdot (\|H_X\| + \|H_O\|) \tag{55a}$$
$$\kappa_3 \cdot (\|H_S\| + \|H_O\|) \le \|\Delta S\| \le \kappa_4 \cdot (\|H_S\| + \|H_O\|). \tag{55b}$$

We only prove the X part; the S part follows directly by symmetry.

1. Since $X'_{\alpha\alpha} - \Lambda_X = (\Delta X)_{\alpha\alpha}$, the first conclusion $X'_{\alpha\alpha} = \Lambda_X + \mathcal{O}(\|\Delta X\|)$ naturally holds.

2. For $X'_{\gamma\alpha}$, we have
$$\|X'_{\gamma\alpha} - \Theta \circ H_O\| \le \kappa_X \cdot \|H_O\| \cdot \|H\|, \tag{56}$$
i.e., there exist $\kappa_5, \kappa_6 > 0$ such that $\kappa_5\|H_O\| \le \|X'_{\gamma\alpha}\| \le \kappa_6\|H_O\|$. From (55), we conclude $X'_{\gamma\alpha} = \mathcal{O}(\min\{\|\Delta X\|, \|\Delta S\|\})$.

3. The norm of $X'_{\gamma\gamma}$ is upper bounded by $\|X'_{\gamma\gamma}\| \le \kappa_X\|H_O\|\|H\|$. On the other hand, we have
$$\|\Delta X\| \cdot \|\Delta S\| \ge \kappa_1\kappa_3 \cdot (\|H_X\| + \|H_O\|) \cdot (\|H_S\| + \|H_O\|) \ge \kappa_7 \cdot \|H_O\| \cdot \|H\|$$
for some positive constant κ_7.

4. Note that

$$
\begin{aligned}
\Lambda_S(\Theta \circ H_O) = (\Lambda_S \Theta) \circ H_O &= \begin{bmatrix} \frac{-\lambda_1 \lambda_{r+1}}{\lambda_1 - \lambda_{r+1}} & \cdots & \frac{-\lambda_r \lambda_{r+1}}{\lambda_r - \lambda_{r+1}} \\ \vdots & \ddots & \vdots \\ \frac{-\lambda_1 \lambda_n}{\lambda_1 - \lambda_n} & \cdots & \frac{-\lambda_r \lambda_n}{\lambda_r - \lambda_n} \end{bmatrix} \circ H_O \\
&= \left((E_{(n-r)\times r} - \Theta) \Lambda_X \right) \circ H_O \\
&= \left((E_{(n-r)\times r} - \Theta) \circ H_O \right) \Lambda_X,
\end{aligned}
$$

where we use the fact that $(AD) \circ B = B \circ (AD) = (B \circ A)D$ for diagonal D. Then,

$$
\begin{aligned}
&\|S'_{\gamma\alpha}\Lambda_X - \Lambda_S X'_{\gamma\alpha}\| \\
&= \|S'_{\gamma\alpha}\Lambda_X - ((E_{(n-r)\times r} - \Theta) \circ H_O)\Lambda_X - (\Lambda_S X'_{\gamma\alpha} - \Lambda_S(\Theta \circ H_O))\| \\
&\leq \|S'_{\gamma\alpha}\Lambda_X - ((E_{(n-r)\times r} - \Theta) \circ H_O)\Lambda_X\| + \|\Lambda_S X'_{\gamma\alpha} - \Lambda_S(\Theta \circ H_O)\| \\
&\leq (\kappa_X + \kappa_S)\kappa_8 \cdot \|H_O\| \cdot \|H\|
\end{aligned}
$$

for some positive constant κ_8. We conclude $S'_{\gamma\alpha}\Lambda_X - \Lambda_S X'_{\gamma\alpha} = \mathcal{O}(\|\Delta X\|\|\Delta S\|)$, following the same argument as in item 3.

□

From the above proof procedure, we see that Theorem 2 provides a subtle and accurate control of the linearization residual, especially the $\|H_O\|$ term. Otherwise, using the classic result in Lemma 1 only, inequalities like (55) and (56) may not be derived in a straightforward manner.

Appendix B Local Linear Convergence with Nondegeneracy but without SC

In this section, we establish the local linear convergence of ADMM applied to SDPs in which primal and dual nondegeneracy hold but strict complementarity fails. That is to say, we consider the case where $Z_\star$ is singular, *i.e.*, $s + r < n$. Again, we assume without loss of generality that $Q_\star = I_n$ in (6). We also need the following index sets

$$
\alpha := \{1, \ldots, r\}, \quad \beta := \{r+1, \ldots, n-s\}, \quad \gamma := \{n-s+1, \ldots, n\},
$$

and then any matrix $H \in \mathbb{S}^n$ can be partitioned as

$$
H = \begin{bmatrix} H_{\alpha\alpha} & H_{\beta\alpha}^\mathsf{T} & H_{\gamma\alpha}^\mathsf{T} \\ H_{\beta\alpha} & H_{\beta\beta} & H_{\gamma\beta}^\mathsf{T} \\ H_{\gamma\alpha} & H_{\gamma\beta} & H_{\gamma\gamma} \end{bmatrix}. \tag{57}
$$

When $Z_\star$ is singular, the projector $\Pi_{\mathbb{S}^n_+}(\cdot)$ is no longer Fréchet differentiable around $Z_\star$ [54, Theorem 4.8]. However, its directional derivative always exists [54, Theorem 4.7].

Lemma 15 ([54, Theorem 4.7]). *The PSD cone projection $\Pi_{\mathbb{S}^n_+}(\cdot)$ is directionally differentiable at $Z_\star$ and, for any $H \in \mathbb{S}^n$ partitioned as in (57), its directional derivative at H is*

$$
\widetilde{\Omega}(H) := \begin{bmatrix} H_{\alpha\alpha} & H_{\beta\alpha}^\mathsf{T} & \widetilde{\Theta}^\mathsf{T} \circ H_{\gamma\alpha}^\mathsf{T} \\ H_{\beta\alpha} & \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ \widetilde{\Theta} \circ H_{\gamma\alpha} & 0 & 0 \end{bmatrix},
$$

where $|\beta| = n - r - s$ is the cardinality of the index set β and the matrix $\widetilde{\Theta} \in \mathbb{R}^{s \times r}$ is defined as

$$
\widetilde{\Theta}_{i,j} := \frac{\lambda_j}{\lambda_j - \lambda_{n-s+i}}, \quad \textit{for } i \in [s],\, j \in [r]. \tag{58}
$$

From the definition of the directional derivative, we have for sufficiently small $H \in \mathbb{S}^n$ that

$$\Pi_{\mathbb{S}^n_+}(Z_\star + H) = \Pi_{\mathbb{S}^n_+}(Z_\star) + \widetilde{\Omega}(H) + o(\|H\|_{\mathsf{F}}). \tag{59}$$

Recall our definition $\Theta^\perp := E_{s\times r} - \Theta$; similarly, we denote $\widetilde{\Omega}^\perp(H)$ as

$$\widetilde{\Omega}^\perp(H) := H - \widetilde{\Omega}(H) = \begin{bmatrix} 0 & 0 & (\Theta^\perp)^\mathsf{T} \circ H_{\gamma\alpha}^\mathsf{T} \\ 0 & -\Pi_{\mathbb{S}^{|\beta|}_+}(-H_{\beta\beta}) & H_{\gamma\beta}^\mathsf{T} \\ \Theta^\perp \circ H_{\gamma\alpha} & H_{\gamma\beta} & H_{\gamma\gamma} \end{bmatrix},$$

where we use the fact that $H_{\beta\beta} = \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) - \Pi_{\mathbb{S}^{|\beta|}_+}(-H_{\beta\beta})$. Similar to (16), we split one-step ADMM operator into two parts:

$$Z^{(k+1)} - Z_\star = \widetilde{\mathcal{M}}(Z^{(k)} - Z_\star) + \widetilde{\Psi}^{(k)},$$

where

$$\begin{aligned} \widetilde{\mathcal{M}}(H) &:= \mathcal{P}\widetilde{\Omega}^\perp(H) + \mathcal{P}^\perp\widetilde{\Omega}(H), & (60)\\ \widetilde{\Psi}^{(k)} &:= (\mathrm{Id} - 2\mathcal{P})(\Pi_{\mathbb{S}^n_+}(Z^{(k)}) - \Pi_{\mathbb{S}^n_+}(Z_\star) - \widetilde{\Omega}(Z^{(k)} - Z_\star)) \\ &= o(\|Z^{(k)} - Z_\star\|_{\mathsf{F}}). & (61) \end{aligned}$$

Although $\widetilde{\mathcal{M}}$ is no longer a linear operator (because of $\Pi_{\mathbb{S}^{|\beta|}_+}$ in $\widetilde{\Omega}$), it is still positively homogenous. This is because $\Pi_{\mathbb{S}^{|\beta|}_+}$ is positively homogenous and other parts of $\widetilde{\mathcal{M}}$ are linear. So we can still obtain the following result similar to Lemma 2.

Lemma 16. *For any matrix $H \in \mathbb{S}^n$ partitioned as in (57), it holds that*

$$\left\langle \widetilde{\Omega}(H), \widetilde{\Omega}^\perp(H) \right\rangle = 2\left\langle \widetilde{\Theta} \circ H_{\gamma\alpha}, \widetilde{\Theta}^\perp \circ H_{\gamma\alpha} \right\rangle \geq 0, \tag{62}$$

with equality if and only if $H_{\gamma\alpha} = 0$, and that

$$\|H\|_{\mathsf{F}}^2 - \|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}}^2 = \|\mathcal{P}\widetilde{\Omega}(H)\|_{\mathsf{F}}^2 + \|\mathcal{P}^\perp\widetilde{\Omega}^\perp(H)\|_{\mathsf{F}}^2 + 4\left\langle \widetilde{\Theta} \circ H_{\gamma\alpha}, \widetilde{\Theta}^\perp \circ H_{\gamma\alpha} \right\rangle. \tag{63}$$

Proof. From the definition of $\widetilde{\Theta}$ (58) and the partition of H (57), we see that

$$\begin{aligned} &\left\langle \widetilde{\Omega}(H), \widetilde{\Omega}^\perp(H) \right\rangle \\ &= \left\langle \begin{bmatrix} H_{\alpha\alpha} & H_{\beta\alpha}^\mathsf{T} & \widetilde{\Theta}^\mathsf{T} \circ H_{\gamma\alpha}^\mathsf{T} \\ H_{\beta\alpha} & \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}) & 0 \\ \widetilde{\Theta} \circ H_{\gamma\alpha} & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 & (\widetilde{\Theta}^\perp)^\mathsf{T} \circ H_{\gamma\alpha}^\mathsf{T} \\ 0 & -\Pi_{\mathbb{S}^{|\beta|}_+}(-H_{\beta\beta}) & H_{\gamma\beta}^\mathsf{T} \\ \widetilde{\Theta}^\perp \circ H_{\gamma\alpha} & H_{\gamma\beta} & H_{\gamma\gamma} \end{bmatrix} \right\rangle \\ &= 2\left\langle \widetilde{\Theta} \circ H_{\gamma\alpha}, \widetilde{\Theta}^\perp \circ H_{\gamma\alpha} \right\rangle \geq 0. \end{aligned} \tag{64}$$

where we already use the fact that $\langle \Pi_{\mathbb{S}^{|\beta|}_+}(H_{\beta\beta}), \Pi_{\mathbb{S}^{|\beta|}_+}(-H_{\beta\beta}) \rangle = 0$. Since all the entries in $\widetilde{\Theta}$ and $\widetilde{\Theta}^\perp$ are strictly positive, the inner product (64) is zero if and only if $H_{\gamma\alpha} = 0$.

To show the second conclusion, we first decompose H as

$$H = \mathcal{P}(\widetilde{\Omega}(H)) + \mathcal{P}(\widetilde{\Omega}^\perp(H)) + \mathcal{P}^\perp(\widetilde{\Omega}(H)) + \mathcal{P}^\perp(\widetilde{\Omega}^\perp(H)).$$

Then, we have

$$\begin{aligned}\|H\|_{\mathsf{F}}^2 &= \|\mathcal{P}(\widetilde{\Omega}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}^{\perp}(\widetilde{\Omega}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}^{\perp}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2 \\ &\quad + 2\left\langle \mathcal{P}(\widetilde{\Omega}(H)), \mathcal{P}(\widetilde{\Omega}^{\perp}(H))\right\rangle + 2\left\langle \mathcal{P}^{\perp}(\widetilde{\Omega}(H)), \mathcal{P}^{\perp}(\widetilde{\Omega}^{\perp}(H))\right\rangle \\ &= \|\mathcal{P}(\widetilde{\Omega}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}^{\perp}(\widetilde{\Omega}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}^{\perp}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2 \\ &\quad + 2\left\langle \widetilde{\Omega}(H), \widetilde{\Omega}^{\perp}(H)\right\rangle,\end{aligned}$$

and

$$\|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}}^2 = \|\mathcal{P}^{\perp}(\widetilde{\Omega}(H)) + \mathcal{P}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2 = \|\mathcal{P}^{\perp}(\widetilde{\Omega}(H))\|_{\mathsf{F}}^2 + \|\mathcal{P}(\widetilde{\Omega}^{\perp}(H))\|_{\mathsf{F}}^2.$$

Combining both expressions with (64) gives the desirable result. □

Theorem 5. *Suppose Assumption 1, primal nondegeneracy (8) and dual nondegeneracy (9) hold. Define*

$$\rho_{\mathrm{ND}} := \sup_{\|H\|_{\mathsf{F}}=1} \|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}} < 1$$

For any $\rho \in (\rho_{\mathrm{ND}}, 1)$, there exists $\bar{k}_{\mathrm{ND}} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}_{\mathrm{ND}}$, it holds that

$$\|Z^{(k+1)} - Z_{\star}\|_{\mathsf{F}} \leq \rho \|Z^{(k)} - Z_{\star}\|_{\mathsf{F}}.$$

Proof. First, we show that for any $H \neq 0$, we have $\|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}} < \|H\|_{\mathsf{F}}$. To see this, suppose there exists a matrix $H \in \mathbb{S}^n$ partitioned as in (57) such that $\|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}} \geq \|H\|_{\mathsf{F}}$. Then, from Lemma 16, we see that

$$\mathcal{P}\widetilde{\Omega}(H) = 0, \quad \mathcal{P}^{\perp}\widetilde{\Omega}^{\perp}(H) = 0, \quad H_{\gamma\alpha} = 0.$$

From $H_{\gamma\alpha} = 0$ condition, we have

$$\begin{aligned}\mathcal{P}\widetilde{\Omega}(H) = 0 \quad &\iff \quad \mathcal{P}\begin{bmatrix} H_{\alpha\alpha} & H_{\beta\alpha}^{\mathsf{T}} & 0 \\ H_{\beta\alpha} & \Pi_{\mathbb{S}_+^{|\beta|}}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} = 0 \\ &\iff \quad \begin{bmatrix} H_{\alpha\alpha} & H_{\beta\alpha}^{\mathsf{T}} & 0 \\ H_{\beta\alpha} & \Pi_{\mathbb{S}_+^{|\beta|}}(H_{\beta\beta}) & 0 \\ 0 & 0 & 0 \end{bmatrix} \in \mathcal{N}(\mathcal{A}) \cap \mathcal{N}_{S_{\star}}.\end{aligned}$$

On the other hand, dual nondegeneracy (9) implies $\mathcal{N}(\mathcal{A}) \cap \mathcal{N}_{S_{\star}} = \{0\}$, and thus

$$H_{\alpha\alpha} = 0, \qquad H_{\beta\alpha} = 0, \qquad \Pi_{\mathbb{S}_+^{|\beta|}}(H_{\beta\beta}) = 0. \tag{65}$$

Symmetrically, we have

$$\begin{aligned}\mathcal{P}^{\perp}\widetilde{\Omega}^{\perp}(H) \quad &\iff \quad \mathcal{P}^{\perp}\begin{bmatrix} 0 & 0 & 0 \\ 0 & -\Pi_{\mathbb{S}_+^{|\beta|}}(-H_{\beta\beta}) & H_{\gamma\beta}^{\mathsf{T}} \\ 0 & H_{\gamma\beta} & H_{\gamma\gamma} \end{bmatrix} = 0 \\ &\iff \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\Pi_{\mathbb{S}_+^{|\beta|}}(-H_{\beta\beta}) & H_{\gamma\beta}^{\mathsf{T}} \\ 0 & H_{\gamma\beta} & H_{\gamma\gamma} \end{bmatrix} \in \mathcal{R}(\mathcal{A}^*) \cap \mathcal{N}_{X_{\star}}.\end{aligned}$$

On the other hand, primal nondegeneracy (8) implies $\mathcal{R}(\mathcal{A}^*) \cap \mathcal{N}_{X_{\star}} = \{0\}$, and thus

$$H_{\gamma\gamma} = 0, \quad H_{\gamma\beta} = 0, \quad \Pi_{\mathbb{S}_+^{|\beta|}}(-H_{\beta\beta}) = 0. \tag{66}$$

Finally, combining (65), (66) together with $H_{\gamma\alpha} = 0$ induces $H = 0$.

Second, we show that $\|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}}/\|H\|_{\mathsf{F}} \leq \rho_{\text{ND}} < 1$ for all $H \neq 0$. Since $\|\widetilde{\mathcal{M}}(\cdot)\|_{\mathsf{F}}$ is continuous and the set $\{H \mid \|H\|_{\mathsf{F}} = 1\}$ is compact, we draw from Extreme Value Theorem that

$$\rho_{\text{ND}} := \sup_{\|H\|_{\mathsf{F}}=1} \|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}} < 1.$$

On the other hand, because $\widetilde{\mathcal{M}}$ is positively homogenous, we have for any $H \neq 0$ that

$$\frac{\|\widetilde{\mathcal{M}}(H)\|_{\mathsf{F}}}{\|H\|_{\mathsf{F}}} = \|\widetilde{\mathcal{M}}(H/\|H\|_{\mathsf{F}})\|_{\mathsf{F}} \leq \sup_{\|H'\|_{\mathsf{F}}=1} \|\widetilde{\mathcal{M}}(H')\|_{\mathsf{F}} < 1.$$

Third, we prove the locally linear decay of $\|Z^{(k)} - Z_\star\|_{\mathsf{F}}$. It follows from (61) that for any $\rho \in (\rho_{\text{ND}}, 1)$, these exists $\bar{k}_{\text{ND}} \in \mathbb{N}$ such that for any $k \geq \bar{k}_{\text{ND}}$, $\|\widetilde{\Psi}^{(k)}\|_{\mathsf{F}} \leq (\rho - \rho_{\text{ND}})\|Z^{(k)} - Z_\star\|_{\mathsf{F}}$. Finally,

$$\begin{aligned}
\|Z^{(k+1)} - Z_\star\|_{\mathsf{F}} &= \|\widetilde{\mathcal{M}}(Z^{(k)} - Z_\star) + \widetilde{\Psi}^{(k)}\|_{\mathsf{F}} \\
&\leq \rho_{\text{ND}} \cdot \|Z^{(k)} - Z_\star\|_{\mathsf{F}} + \|\widetilde{\Psi}^{(k)}\|_{\mathsf{F}} \\
&\leq (\rho_{\text{ND}} + \rho - \rho_{\text{ND}}) \cdot \|Z^{(k)} - Z_\star\|_{\mathsf{F}} \\
&= \rho\|Z^{(k)} - Z_\star\|_{\mathsf{F}}.
\end{aligned}$$

□

Appendix C Missing Materials in Section 6

C.1 Proof of Lemma 4

From one-step ADMM (4), we have

$$\begin{aligned}
Z^{(k+1)} - Z^{(k)} &= -2\mathcal{P}\Pi_{\mathbb{S}_+^n}(Z^{(k)}) + \mathcal{P}Z^{(k)} + \Pi_{\mathbb{S}_+^n}(Z^{(k)}) + \mathcal{A}^\dagger b - \sigma\mathcal{P}^\perp C - Z^{(k)} \\
&= -2\mathcal{P}X^{(k)} + X^{(k)} - \mathcal{P}^\perp Z^{(k)} + \mathcal{A}^\dagger b - \sigma\mathcal{P}^\perp C \\
&= -\mathcal{P}X^{(k)} + \mathcal{P}^\perp X^{(k)} - \mathcal{P}^\perp(X^{(k)} - \sigma S^{(k)}) + \mathcal{A}^\dagger b - \sigma\mathcal{P}^\perp C \\
&= -\mathcal{P}X^{(k)} + \mathcal{A}^\dagger b + \sigma\mathcal{P}^\perp(S^{(k)} - C).
\end{aligned}$$

Since for any $\widetilde{X}$ such that $\mathcal{A}\widetilde{X} = b$,

$$\mathcal{P}\widetilde{X} = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}\widetilde{X} = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}b = \mathcal{A}^\dagger b,$$

the equality (25) follows from the orthogonality between $\mathcal{P}$ and $\mathcal{P}^\perp$.

Now we prove the inequality (26). Let $\overline{Z}$ be an arbitrary point in $\mathcal{Z}_\star$; *i.e.*, $\overline{Z}$ may not be the convergent point $Z_\star$ of ADMM. Define $\overline{X} := \Pi_{\mathbb{S}_+^n}(\overline{Z})$ and $\overline{S} := (1/\sigma)\Pi_{\mathbb{S}_+^n}(-\overline{Z})$. So,

$$\mathcal{P}(\Pi_{\mathbb{S}_+^n}(\overline{Z}) - \widetilde{X}) = 0, \qquad \mathcal{P}^\perp(\Pi_{\mathbb{S}_+^n}(\overline{Z}) - C) = 0$$

for any matrix $\widetilde{X} \in \mathbb{S}^n$ satisfying $\mathcal{A}\widetilde{X} = b$. Then, we have

$$\begin{aligned}
&\|Z^{(k)} - \overline{Z}\|_{\mathsf{F}}^2 - \|Z^{(k+1)} - \overline{Z}\|_{\mathsf{F}}^2 \\
= &\|Z^{(k)} - \overline{Z}\|_{\mathsf{F}}^2 - \|Z^{(k+1)} - Z^{(k)} + Z^{(k)} - \overline{Z}\|_{\mathsf{F}}^2 \\
= &-2\langle Z^{(k+1)} - Z^{(k)}, Z^{(k)} - \overline{Z}\rangle - \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 \\
= &2\langle \mathcal{P}(X^{(k)} - \overline{X}) - \sigma\mathcal{P}^\perp(S^{(k)} - \overline{S}), Z^{(k)} - Z_\star\rangle - \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2. \qquad (67)
\end{aligned}$$

We further decompose $Z^{(k)} - \overline{Z}$ as

$$
\begin{aligned}
& Z^{(k)} - \overline{Z} \\
&= X^{(k)} - \sigma S^{(k)} - (\overline{X} - \sigma \overline{S}) \\
&= \mathcal{P}(X^{(k)} - \overline{X}) + \mathcal{P}^{\perp}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}^{\perp}(S^{(k)} - C) - \sigma \mathcal{P}^{\perp}(S^{(k)} - C).
\end{aligned}
$$

Then the inner product term on the right-hand side of (67) becomes

$$
\begin{aligned}
& \langle \mathcal{P}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}^{\perp}(S^{(k)} - \overline{S}), Z^{(k)} - Z_\star \rangle \\
&= \langle \mathcal{P}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}^{\perp}(S^{(k)} - \overline{S}), \mathcal{P}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}^{\perp}(S^{(k)} - C) \rangle \\
&\quad - \langle \mathcal{P}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}^{\perp}(S^{(k)} - \overline{S}), \mathcal{P}^{\perp}(X^{(k)} - \overline{X}) - \sigma \mathcal{P}(S^{(k)} - C) \rangle \\
&= \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 - \sigma \langle \mathcal{P}(X^{(k)} - \overline{X}), \mathcal{P}^{\perp}(\mathcal{P}(S^{(k)} - \overline{S})) \\
&\quad - \sigma \langle \mathcal{P}^{\perp}(X^{(k)} - \overline{X}), \mathcal{P}^{\perp}(S^{(k)} - \overline{S}) \rangle \\
&= \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 - \sigma \langle X^{(k)} - \overline{X}, S^{(k)} - \overline{S} \rangle \\
&= \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 - \sigma \langle X^{(k)}, \overline{S} \rangle + \sigma \langle \overline{X}, S^{(k)} \rangle \\
&\geq \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2,
\end{aligned}
$$

where the last equality uses the fact that $\langle \overline{X}, \overline{S} \rangle = 0$ and $\langle X^{(k)}, S^{(k)} \rangle = 0$. Combining with (67) yields

$$
\|Z^{(k)} - \overline{Z}\|_{\mathsf{F}}^2 - \|Z^{(k+1)} - \overline{Z}\|_{\mathsf{F}}^2 \geq \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2.
$$

Now we choose $\overline{Z}$ as the closest point in $\mathcal{Z}_\star$ to $Z^{(k)}$. Then, we have

$$
\begin{aligned}
\operatorname{dist}^2(Z^{(k)}, \mathcal{Z}_\star) &= \|Z^{(k)} - \overline{Z}\|_{\mathsf{F}}^2 \\
&\geq \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 + \|Z^{(k+1)} - \overline{Z}\|_{\mathsf{F}}^2 \\
&\geq \|Z^{(k+1)} - Z^{(k)}\|_{\mathsf{F}}^2 + \operatorname{dist}^2(Z^{(k+1)}, \mathcal{Z}_\star).
\end{aligned}
$$

C.2 Proof of Lemma 7

(1) First, we show that $(X_\star, S_\star)$ is a KKT point for (1) if and only if

$$
\mathcal{P}(X_\star - \widetilde{X}) = 0, \;\; \mathcal{P}^{\perp}(S_\star - C) = 0, \;\; \langle X_\star, C \rangle + \langle \widetilde{X}, S_\star \rangle - \langle \widetilde{X}, C \rangle = 0, \;\;\; X \in \mathbb{S}_+^n, \;\;\; S \in \mathbb{S}_+^n,
$$

where $\widetilde{X} \in \mathbb{S}^n$ is an arbitrary matrix satisfying $\mathcal{A}\widetilde{X} = b$.

- If $\mathcal{A}X_\star = b$, then $\mathcal{P}(X_\star - \widetilde{X}) = \mathcal{A}^{\dagger}\mathcal{A}(X_\star - \widetilde{X}) = \mathcal{A}^{\dagger}(b - b) = 0$. The converse holds because $\mathcal{A}$ is surjective and thus $\mathcal{A}\mathcal{A}^{\dagger} = \mathrm{Id}$. Together with $X_\star \in \mathbb{S}_+^n$, it gives primal feasibility.
- Similarly, we note that $\mathcal{P}^{\perp}(S_\star - C) = 0$ is equivalent to $S_\star - C \in \mathcal{R}(\mathcal{A}^*)$, which is further equivalent to $\mathcal{A}^* y + S = C$ for some $y \in \mathbb{R}^m$. Together with $S_\star \in \mathbb{S}_+^n$, this gives dual feasibility.
- The third condition implies zero duality gap:

$$
\begin{aligned}
\langle X_\star, C \rangle - b^{\mathsf{T}} y_\star = 0 \quad &\Longleftrightarrow \quad \langle X_\star, C \rangle + \left\langle b, (\mathcal{A}^*)^{\dagger}(S_\star - C) \right\rangle = 0 \\
&\Longleftrightarrow \quad \langle X_\star, C \rangle + \langle \mathcal{A}^{\dagger} b, S_\star \rangle - \langle \mathcal{A}^{\dagger} b, C \rangle = 0 \\
&\Longleftrightarrow \quad \langle X_\star, C \rangle + \langle \widetilde{X}, S_\star \rangle - \langle \widetilde{X}, C \rangle = 0.
\end{aligned}
$$

For notational convenience, we define

$$
\begin{aligned}
\mathcal{F}_X \; &:= \{(X, \sigma S) \mid \mathcal{P}(X - \widetilde{X}) = 0\}, \qquad & \mathcal{F}_Z \; &:= \{(X, \sigma S) \mid \mathcal{P}^{\perp}(S - C) = 0\}, \\
\mathcal{R}_X \; &:= \{(X, \sigma S) \mid \Pi_{\mathcal{T}_{S_\star}}(X) = 0\}, \qquad & \mathcal{R}_S \; &:= \{(X, \sigma S) \mid \Pi_{\mathcal{T}_{X_\star}}(\sigma S) = 0\},
\end{aligned}
$$

$$
\mathcal{F}_{\mathrm{gap}} := \{(X, \sigma S) \mid \langle X_\star, \sigma C \rangle + \langle \widetilde{X}, \sigma S_\star \rangle - \langle \widetilde{X}, \sigma C \rangle = 0\},
$$

$$
\mathcal{F} := \mathcal{F}_X \cap \mathcal{F}_S \cap \mathcal{F}_{\mathrm{gap}}, \qquad \mathcal{R} := \mathcal{R}_X \cap \mathcal{R}_S.
$$

(Note that the choice of $\widetilde{X}$ does not matter in $\mathcal{F}_X$, as long as $\mathcal{A}\widetilde{X} = b$.)

(2) Second, we show that $\mathcal{X}_\star \cap (\sigma\mathcal{S}_\star) = \mathcal{F} \cap (\mathbb{S}^n_+ \times \mathbb{S}^n_+) = \mathcal{F} \cap (\mathbb{S}^n_+ \times \mathbb{S}^n_+) \times \mathcal{R}$. To see this, we choose any pair of optimal solutions $(X, \sigma S) \in \mathcal{X}_\star \times (\sigma\mathcal{S}_\star)$ and write down the complementary slackness condition: $\langle X, \sigma S_\star\rangle = 0$ and $\langle X_\star, \sigma S\rangle = 0$. Combined with the facts that $X \in \mathbb{S}^n_+$ and $\sigma S \in \mathbb{S}^n_+$, we have

$$X \in \operatorname{minface}\left(X_\star, \mathbb{S}^n_+\right) = \mathbb{S}^n_+ \times \mathcal{N}_{S_\star}, \qquad \sigma S \in \operatorname{minface}\left(\sigma S_\star, \mathbb{S}^n_+\right) = \mathbb{S}^n_+ \cap \mathcal{N}_{X_\star}.$$

Thus, $\Pi_{\mathcal{T}_{S_\star}}(X) = 0$ and $\Pi_{\mathcal{T}_{X_\star}}(\sigma S) = 0$, which directly implies $(X, \sigma S) \in \mathcal{R}_X \cap \mathcal{R}_S = \mathcal{R}$.

(3) Third, we project an arbitrary $(X, \sigma S) \in \mathbb{S}^n \times \mathbb{S}^n$ to the regularized linear system $\mathcal{F} \cap \mathcal{R}$. By Hoffman's error bound [26], there exists $\kappa_0 > 0$ such that

$$\begin{aligned}
&\kappa_0 \cdot \operatorname{dist}((X, \sigma S), \mathcal{F} \cap \mathcal{R}) \\
\le\ &\operatorname{dist}((X, \sigma S), \mathcal{F}_X) + \operatorname{dist}((X, \sigma S), \mathcal{F}_S) + \operatorname{dist}((X, \sigma S), \mathcal{F}_{\text{gap}}) \\
&+ \operatorname{dist}((X, \sigma S), \mathcal{R}_X) + \operatorname{dist}((X, \sigma S), \mathcal{R}_S),
\end{aligned} \tag{68}$$

where

$$\begin{aligned}
\operatorname{dist}((X, \sigma S), \mathcal{F}_X) &= \|\mathcal{P}(X - \widetilde{X})\|_{\mathsf{F}}, & \operatorname{dist}((X, \sigma S), \mathcal{F}_Z) &= \sigma\|\mathcal{P}^\perp(S - C)\|_{\mathsf{F}} \\
\operatorname{dist}((X, \sigma S), \mathcal{R}_X) &= \|\Pi_{\mathcal{T}_{S_\star}}(X)\|_{\mathsf{F}}, & \operatorname{dist}((X, \sigma S), \mathcal{R}_S) &= \sigma\|\Pi_{\mathcal{T}_{X_\star}}(S)\|_{\mathsf{F}},
\end{aligned}$$
$$\operatorname{dist}((X, \sigma S), \mathcal{F}_{\text{gap}}) = \frac{\sigma}{\sqrt{\sigma^2\|C\|_{\mathsf{F}}^2 + \|\mathcal{A}^\dagger b\|_{\mathsf{F}}^2}} |\langle X_\star, C\rangle + \langle \widetilde{X}, S_\star\rangle - \langle \widetilde{X}, C\rangle|.$$

(4) Fourth, we explicitly construct a point belonging to $\mathcal{X}_\star \times (\sigma\mathcal{S}_\star)$. Take

$$\beta := \max\left\{\frac{1}{\lambda_r} \cdot \left([-\lambda_{\min}(X)]_+ + \|Z_X\|_2\right),\ \frac{1}{-\lambda_{r+1}} \cdot \left([-\lambda_{\min}(\sigma S)]_+ + \|Z_S\|_2\right)\right\}$$

and define $(Z_X, Z_S) := (X, \sigma S) - \Pi_{\mathcal{F}\cap\mathcal{R}}(X, \sigma S)$. Thus, by definition

$$\sqrt{\|Z_X\|_{\mathsf{F}}^2 + \|Z_S\|_{\mathsf{F}}^2} = \operatorname{dist}((X, \sigma S), \mathcal{F} \cap \mathcal{R}).$$

Then, consider the point $(X, \sigma S) - (Z_X, Z_S) + \beta \cdot (X_\star, \sigma S_\star)$. For the primal part:

$$\begin{aligned}
&\lambda_{\min}\left((X - Z_X + \beta \cdot X_\star)_{1:r,1:r}\right) \\
\ge\ &\beta \cdot \lambda_{\min}\left((X_\star)_{1:r,1:r}\right) + \lambda_{\min}\left((X - Z_X)_{1:r,1:r}\right) && \text{(69a)} \\
=\ &\beta\lambda_r + \min\left\{\lambda_{\min}(X - Z_X), 0\right\} && \text{(69b)} \\
\ge\ &\beta\lambda_r + \min\left\{\lambda_{\min}(X) - \|Z_X\|_2, 0\right\} && \text{(69c)} \\
\ge\ &\frac{1}{\lambda_r}\left([-\lambda_{\min}(X)]_+ + \|Z_X\|_2\right) \cdot \lambda_r + \min\left\{\lambda_{\min}(X) - \|Z_X\|_2, 0\right\} \\
\ge\ &0,
\end{aligned}$$

where (69a) and (69c) come form Weyl's inequality, (69b) holds since $X - Z_X \in \mathcal{N}_{S_\star}$. Combining the above inequality with the facts taht $X_\star \in \mathcal{N}_{S_\star}$ and $X - Z_X \in \mathcal{N}_{S_\star}$ gives

$$\lambda_{\min}(X - Z_X + \beta X_\star) = \min\left\{\lambda_{\min}\left((X - Z_X + \beta X_\star)_{1:r,1:r}\right), 0\right\} \ge 0.$$

Symmetrically, $\lambda_{\min}(\sigma S - Z_S + \beta\sigma S_\star) \ge 0$. Therefore, $(X - Z_X + \beta X_\star, \sigma S - Z_S + \beta\sigma S_\star) \in \mathbb{S}^n_+ \times \mathbb{S}^n_+$. Combining the fact that both $(X - Z_X, \sigma S - Z_S)$ and $(X_\star, \sigma S_\star)$ belong to $\mathcal{F} \cap \mathcal{R}$, we conclude that

$$\frac{1}{1+\beta} \cdot (X - Z_X + \beta X_\star, \sigma S - Z_S + \beta\sigma S_\star) \in \mathcal{F} \cap \mathcal{R} \cap (\mathbb{S}^n_+ \times \mathbb{S}^n_+) = \mathcal{X}_\star \cap (\sigma\mathcal{S}_\star).$$

(5) Finally, we upper bound the distance from $(X, \sigma S)$ to $\mathcal{X}_\star \cap (\sigma \mathcal{S}_\star)$ by

$$
\begin{aligned}
&\operatorname{dist}((X, \sigma S), \mathcal{X}_\star \times \sigma \mathcal{S}_\star) \\
\leq\ & \left\| (X, \sigma S) - \frac{1}{1+\beta} \cdot (X - Z_X + \beta X_\star, \sigma S - Z_S + \beta \sigma S_\star) \right\|_{\mathsf{F} \times \mathsf{F}} \\
=\ & \sqrt{\left\| X - \frac{1}{1+\beta}(X - Z_X + \beta X_\star) \right\|_{\mathsf{F}}^2 + \left\| \sigma S - \frac{1}{1+\beta}(\sigma S - Z_S + \beta \sigma S_\star) \right\|_{\mathsf{F}}^2} \\
=\ & \sqrt{2} \left\| X - \frac{1}{1+\beta}(X - Z_X + \beta X_\star) \right\|_{\mathsf{F}} + \sqrt{2} \left\| \sigma S - \frac{1}{1+\beta}(\sigma S - Z_S + \beta \sigma S_\star) \right\|_{\mathsf{F}}.
\end{aligned} \tag{70}
$$

To bound the first term on the right-hand side, we have

$$
\begin{aligned}
\left\| X - \frac{1}{1+\beta} \cdot (X - Z_X + \beta X_\star) \right\|_{\mathsf{F}} &= \left\| \frac{1}{1+\beta} \cdot (\beta \cdot X - Z_X + \beta X_\star) \right\|_{\mathsf{F}} \\
&\leq \frac{1}{1+\beta} \cdot (\|\beta X\|_{\mathsf{F}} + \|Z_X\|_{\mathsf{F}} + \|\beta X_\star\|_{\mathsf{F}}) \\
&\leq \beta \|X\|_{\mathsf{F}} + \|Z_X\|_{\mathsf{F}} + \beta \|X_\star\|_{\mathsf{F}} \\
&\leq (\delta_X + \lambda_1)\beta + \|Z_X\|_{\mathsf{F}}.
\end{aligned} \tag{71}
$$

Similarly, we have

$$
\left\| \sigma S - \frac{1}{1+\beta} \cdot (\sigma S - Z_S + \beta \sigma S_\star) \right\|_{\mathsf{F}} \leq (\delta_S - \lambda_n)\beta + \|Z_S\|_{\mathsf{F}}. \tag{72}
$$

Combining (70)–(72) gives

$$
\begin{aligned}
&\operatorname{dist}((X, \sigma S), \mathcal{X}_\star \times (\sigma \mathcal{S}_\star)) \\
\leq\ & \|Z_X\|_{\mathsf{F}} + \|Z_S\|_{\mathsf{F}} + (\delta_X + \delta_S + \lambda_1 - \lambda_n)\beta \\
\leq\ & \sqrt{2(\|Z_X\|_{\mathsf{F}}^2 + \|Z_S\|_{\mathsf{F}}^2)} + (\delta_X + \delta_S + \lambda_1 - \lambda_n) \\
& \cdot \max\left\{ \frac{1}{\lambda_r} \cdot ([-\lambda_{\min}(X)]_+ + \|Z_X\|_2),\ \frac{1}{-\lambda_{r+1}} \cdot ([-\lambda_{\min}(\sigma S)]_+ + \|Z_S\|_2) \right\} \\
\leq\ & \sqrt{2(\|Z_X\|_{\mathsf{F}}^2 + \|Z_S\|_{\mathsf{F}}^2)} + (\delta_X + \delta_S + \lambda_1 - \lambda_n) \cdot \max\left\{ \frac{1}{\lambda_r}, \frac{1}{-\lambda_{r+1}} \right\} \\
& \cdot ([-\lambda_{\min}(X)]_+ + \|Z_X\|_2 + [-\lambda_{\min}(\sigma S)]_+ + \|Z_S\|_2) \\
\leq\ & \kappa_1 \cdot ([-\lambda_{\min}(X)]_+ + [-\lambda_{\min}(\sigma S)]_+) + \kappa_2 \cdot \sqrt{\|Z_X\|_{\mathsf{F}}^2 + \|Z_S\|_{\mathsf{F}}^2} \\
\leq\ & \kappa_1 \cdot ([-\lambda_{\min}(X)]_+ + [-\lambda_{\min}(\sigma S)]_+) + \kappa_2 \cdot \operatorname{dist}((X, \sigma S), \mathcal{F} \cap \mathcal{R}),
\end{aligned} \tag{73}
$$

where

$$
\kappa_1 := (\delta_X + \delta_S + \lambda_1 - \lambda_n) \cdot \max\left\{ \frac{1}{\lambda_r}, \frac{1}{-\lambda_{r+1}} \right\}, \quad \kappa_2 := \kappa_1 + \sqrt{2}.
$$

Therefore, combining (68) and (73) yields the desirable result with

$$
\kappa = \left(\kappa_1 + \frac{\kappa_2}{\kappa_0} \cdot \left(1 + \frac{1}{\sqrt{\sigma^2 \|C\|_{\mathsf{F}}^2 + \|\mathcal{A}^\dagger b\|_{\mathsf{F}}^2}} \right) \right)^{-1}.
$$

Appendix D Missing Materials in Section 7

D.1 Proof of Lemma 9

The proof of Lemma 9 needs the following two auxiliary lemmas.

Lemma 17. *Let $S \in \mathbb{S}^n$ satisfy $\|S\|_2 \le \frac{3}{4}$ and define $\psi(S) := \exp(S) - I - S$. Then, it holds that*

$$\|\psi(S)\|_2 \le \frac{2}{3}\|S\|_2^2.$$

Proof. From the definition of ψ, we have

$$\psi(S) = \exp(S) - I - S = \sum_{k=2}^{\infty} \frac{1}{k!} S^k,$$

and then

$$\begin{aligned}\|\psi(S)\|_2 &\le \sum_{k=2}^{\infty} \frac{1}{k!}\|S\|_2^k = \frac{1}{2}\|S\|_2^2 \cdot \sum_{k=0}^{\infty} \frac{2}{(k+2)!}\|S\|_2^k \le \frac{1}{2}\|S\|_2^2 \cdot \sum_{k=0}^{\infty} \left(\frac{1}{3}\right)^k \|S\|_2^k \\ &= \frac{1}{2}\|S\|_2^2 \cdot \frac{1}{1 - \frac{1}{3}\|S\|_2},\end{aligned}$$

where the last inequality uses the fact that $\frac{2}{n!} \le \left(\frac{1}{3}\right)^{n-2}$ for all $n \ge 3$. Finally, we conclude from the assumption that $\|S\|_2 \le \frac{3}{4}$:

$$\|\psi(S)\|_2 \le \frac{1}{2}\|S\|_2^2 \cdot \frac{4}{3} = \frac{2}{3}\|S\|_2^2.$$

□

Lemma 18. *Let $X \in \mathbb{S}^n$ be partitioned as*

$$X = \begin{bmatrix} A & B^\top \\ B & C \end{bmatrix} \in \mathbb{S}^n.$$

Then, it holds that

$$\max\{\|A\|_2, \|B\|_2, \|C\|_2\} \le \|X\|_2 \le \|A\|_2 + \|B\|_2 + \|C\|_2.$$

Proof. On one hand, we have

$$\|X\|_2 = \sup_{\|x\|_2^2 + \|y\|_2^2 = 1} \left\| \begin{bmatrix} A & B^\top \\ B & C \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \right\|_2 \ge \sup_{\|x\|_2^2 = 1} \left\| \begin{bmatrix} A & B^\top \\ B & C \end{bmatrix} \begin{bmatrix} x \\ 0 \end{bmatrix} \right\|_2 = \sup_{\|x\|_2^2 = 1} \left\| \begin{bmatrix} Ax \\ Bx \end{bmatrix} \right\|_2,$$

and then

$$\|X\|_2 \ge \sup_{\|x\|_2^2=1} \|Ax\|_2 = \|A\|_2, \qquad \|X\|_2 \ge \sup_{\|x\|_2^2=1} \|Bx\|_2 = \|B\|_2.$$

Similarly, $\|X\|_2 \le \|C\|_2$.

On the other hand, we see that

$$\begin{aligned}\|X\|_2 &\le \left\| \begin{bmatrix} A & 0 \\ 0 & 0 \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} 0 & 0 \\ 0 & C \end{bmatrix} \right\|_2 + \left\| \begin{bmatrix} 0 & B^\top \\ B & 0 \end{bmatrix} \right\|_2 \\ &= \|A\|_2 + \|C\|_2 + \left\| \begin{bmatrix} 0 & B^\top \\ B & 0 \end{bmatrix} \right\|_2 \\ &= \|A\|_2 + \|C\|_2 + \|B\|_2.\end{aligned}$$

□

Now we are ready to prove Lemma 9.

Proof of Lemma 9. Since $Z_{X,\ell} \in \mathbb{S}^n_{++}$ and $Z_{S,\ell} \in \mathbb{S}^n_{--}$, the Sylvester equation (36) has a unique solution with

$$\begin{aligned}\operatorname{vec}(W_{O,\ell}) &= (I_r \otimes (-Z_{S,\ell}) + Z_{X,\ell} \otimes I_{n-r})^{-1} \operatorname{vec}(Z_{O,\ell}) \\ &= ((-Z_{S,\ell}) \oplus Z_{X,\ell})^{-1} \operatorname{vec}(Z_{O,\ell})\end{aligned}$$

From [52, Theorem 2.5], we see that the eigenvalues of $(-Z_{S,\ell}) \oplus Z_{X,\ell}$ equal to the sum of the eigenvalues of $-Z_{S,\ell}$ and $Z_{X,\ell}$. Thus, $(-Z_{S,\ell}) \oplus Z_{X,\ell}$ is positive definite and

$$\begin{aligned}\|\operatorname{vec}(W_{O,\ell})\|_2 &\le \frac{1}{\lambda_{\min}\left((-Z_{S,\ell}) \oplus Z_{X,\ell}\right)} \cdot \|\operatorname{vec}(Z_{O,\ell})\|_2 \\ &= \frac{1}{\lambda_{\min}\left(Z_{X,\ell}\right) - \lambda_{\max}\left(Z_{S,\ell}\right)} \cdot \|\operatorname{vec}(Z_{O,\ell})\|_2.\end{aligned}$$

Therefore, we can upper bound $\|W_{O,\ell}\|_2$ and $\|W_\ell\|_2$ by $\|Z_{O,\ell}\|_2$:

$$\|W_{O,\ell}\|_2 \le \|W_{O,\ell}\|_{\mathsf{F}} = \|\operatorname{vec}(W_{O,\ell})\|_2 \le \frac{d}{\lambda_{\min}\left(Z_{X,\ell}\right) - \lambda_{\max}\left(Z_{S,\ell}\right)} \cdot \|Z_{O,\ell}\|_2 = \eta_\ell \cdot \|Z_{O,\ell}\|_2 \tag{74}$$

and

$$\|W_\ell\|_2 = \left\| \begin{bmatrix} 0 & -W_{O,\ell}^\mathsf{T} \\ W_{O,\ell} & 0 \end{bmatrix} \right\|_2 = \|W_{O,\ell}\|_2 \le \eta_\ell \cdot \|Z_{O,\ell}\|_2 \tag{75}$$

In addition, we have

$$\begin{bmatrix} I_r & W_{O,\ell}^\mathsf{T} \\ -W_{O,\ell} & I_{n-r} \end{bmatrix} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} \begin{bmatrix} I_r & -W_{O,\ell}^\mathsf{T} \\ W_{O,\ell} & I_{n-r} \end{bmatrix} = \begin{bmatrix} Q_X & Q_O^\mathsf{T} \\ Q_O & Q_S \end{bmatrix}, \quad \text{where } W_\ell = \begin{bmatrix} 0 & -W_{O,\ell}^\mathsf{T} \\ W_{O,\ell} & 0 \end{bmatrix}$$

and

$$\begin{aligned} Q_X &= Z_{X,\ell} + W_{O,\ell}^\mathsf{T} Z_{O,\ell} + Z_{O,\ell}^\mathsf{T} W_{O,\ell} + W_{O,\ell}^\mathsf{T} Z_{S,\ell} W_{O,\ell} \\ Q_S &= W_{O,\ell} Z_{X,\ell} W_{O,\ell}^\mathsf{T} - Z_{O,\ell} W_{O,\ell}^\mathsf{T} - W_{O,\ell} Z_{O,\ell}^\mathsf{T} + Z_{S,\ell} \\ Q_O &= -W_{O,\ell} Z_{X,\ell} + Z_{O,\ell} - W_{O,\ell} Z_{O,\ell}^\mathsf{T} W_{O,\ell} + Z_{S,\ell} W_{O,\ell}. \end{aligned}$$

From the definition of $W_{O,\ell}$ (36), we further have $Q_O = -W_{O,\ell} Z_{O,\ell}^\mathsf{T} W_{O,\ell}$.

Now, we are ready to bound the following spectral norm:

$$\begin{aligned} &\left\| \begin{bmatrix} Z_{X,\ell+1} - Z_{X,\ell} & Z_{O,\ell+1}^\mathsf{T} \\ Z_{O,\ell+1} & Z_{S,\ell+1} - Z_{S,\ell} \end{bmatrix} \right\|_2 \\ = &\left\| \exp(W_\ell)^\mathsf{T} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} \exp(W_\ell) - \begin{bmatrix} Z_{X,\ell} & 0 \\ 0 & Z_{S,\ell} \end{bmatrix} \right\|_2 \\ \le &\left\| (I + W_\ell)^\mathsf{T} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} (I + W_\ell) - \begin{bmatrix} Z_{X,\ell} & 0 \\ 0 & Z_{S,\ell} \end{bmatrix} \right\|_2 \\ &+ \left\| \exp(W_\ell)^\mathsf{T} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} \exp(W_\ell) - (I + W_\ell)^\mathsf{T} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} (I + W_\ell) \right\|_2. \end{aligned} \tag{76}$$

We then bound the two terms on right-hand side of (76) one-by-one.

1. By definition, we have

$$
\begin{aligned}
&\left\| (I+W_\ell)^\mathsf{T} \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\mathsf{T} \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} (I+W_\ell) - \begin{bmatrix} Z_{X,\ell} & 0 \\ 0 & Z_{S,\ell} \end{bmatrix} \right\|_2 \\
&= \left\| \begin{bmatrix} Q_X - Z_{X,\ell} & Q_O^\mathsf{T} \\ Q_O & Q_S - Q_{S,\ell} \end{bmatrix} \right\|_2 \\
&\le \|Q_S - Z_{X,\ell}\|_2 + \|Q_O\|_2 + \|Q_S - Z_{S,\ell}\|_2 \\
&= \|W_{O,\ell}^\mathsf{T} Z_{O,\ell} + Z_{O,\ell}^\mathsf{T} W_{O,\ell} + W_{O,\ell}^\mathsf{T} Z_{S,\ell} W_{O,\ell}\|_2 + \|W_{O,\ell} Z_{O,\ell}^\mathsf{T} W_{O,\ell}\|_2 \\
&\quad + \|W_{O,\ell} Z_{X,\ell} W_{O,\ell}^\mathsf{T} - Z_{O,\ell} W_{O,\ell}^\mathsf{T} - W_{O,\ell} Z_{O,\ell}^\mathsf{T}\|_2,
\end{aligned} \tag{77}
$$

where the inequality follows from Lemma 18. We can further bound the three terms on the right-hand side of (77) one-by-one.

(a) For the first term, we have from (74) that

$$
\begin{aligned}
&\|W_{O,\ell}^\mathsf{T} Z_{O,\ell} + Z_{O,\ell}^\mathsf{T} W_{O,\ell} + W_{O,\ell}^\mathsf{T} Z_{S,\ell} W_{O,\ell}\|_2 \\
&\le 2\|W_{O,\ell}\|_2 \cdot \|Z_{O,\ell}\|_2 + \|W_{O,\ell}\|_2^2 \cdot \|Z_{S,\ell}\|_2 \\
&\le 2\eta_\ell \cdot \|Z_{O,\ell}\|_2^2 + \eta_\ell^2 \cdot \|Z_{S,\ell}\|_2 \cdot \|Z_{O,\ell}\|_2^2 \\
&\le 2\eta_\ell \cdot \|Z_{O,\ell}\|_2^2 + \eta_\ell^2 \cdot \|Z_\ell\|_2 \cdot \|Z_{O,\ell}\|_2^2.
\end{aligned} \tag{78}
$$

(b) For the second term, we obtain from (74) that

$$
\|W_{O,\ell} Z_{O,\ell}^\mathsf{T} W_{O,\ell}\|_2 \le \eta_\ell^2 \cdot \|Z_{O,\ell}\|_2^3 \le \eta_\ell^2 \cdot \|Z_\ell\|_2 \cdot \|Z_{O,\ell}\|_2^2. \tag{79}
$$

(c) Again from (74), we have

$$
\|W_{O,\ell} Z_{X,\ell} W_{O,\ell}^\mathsf{T} - Z_{O,\ell} W_{O,\ell}^\mathsf{T} - W_{O,\ell} Z_{O,\ell}^\mathsf{T}\|_2 \le 2\eta_\ell \cdot \|Z_{O,\ell}\|_2^2 + \eta_\ell^2 \cdot \|Z_\ell\|_2 \cdot \|Z_{O,\ell}\|_2^2. \tag{80}
$$

2. Similarly, the second term on the right-hand side of (76) can be bounded as

$$
\begin{aligned}
&\|\exp(W_\ell)^\mathsf{T} Z_\ell \exp(W_\ell) - (I+W_\ell)^\mathsf{T} Z_\ell (I+W_\ell)\|_2 \\
&\le 2\|\exp(W_\ell) - I_n - W_\ell\|_2 \cdot \|Z_\ell\|_2 \cdot \|I_n + W_\ell\|_2 + \|\exp(W_\ell) - I_n - W_\ell\|_2^2 \cdot \|Z_\ell\|_2.
\end{aligned} \tag{81}
$$

Again, we bound the two terms on the right-hand side one-by-one.

(a) We see from Lemma 17 and (75) that

$$
\begin{aligned}
&\|\exp(W_\ell) - I_n - W_\ell\|_2 \cdot \|Z_\ell\|_2 \cdot \|I_n + W_\ell\|_2 \\
&\le \frac{2}{3} \cdot \|W_\ell\|_2^2 \cdot \|Z_\ell\|_2 \cdot \|I_n + W_\ell\|_2 \\
&\le \frac{2}{3} \cdot \|W_\ell\|_2^2 \cdot \|Z_\ell\|_2 \cdot (1 + \|W_\ell\|_2) \\
&\le \frac{2}{3}\eta_\ell^2 \cdot \|Z_{O,\ell}\|_2^2 \cdot \|Z_\ell\|_2 \cdot (1 + \eta_\ell \|Z_{O,\ell}\|_2) \\
&\le \frac{2}{3}\eta_\ell^2 \cdot \|Z_\ell\|_2 \cdot \|Z_{O,\ell}\|_2^2 + \frac{2}{3}\eta_\ell^3 \cdot \|Z_\ell\|_2^2 \cdot \|Z_{O,\ell}\|_2^2.
\end{aligned} \tag{82}
$$

(b) The second term on the right-hand side of (81) can be readily bounded by

$$
\|\exp(W_\ell) - I_n - W_\ell\|_2^2 \cdot \|Z_\ell\|_2 \le \frac{4}{9} \cdot \|W_\ell\|_2^4 \cdot \|Z_\ell\|_2 \le \frac{4}{9}\eta_\ell^4 \cdot \|Z_\ell\|_2^3 \cdot \|Z_{O,\ell}\|_2^2. \tag{83}
$$

Then, combining Lemma 18 and (76)–(83) yields

$$
\begin{aligned}
&\max\{\|Z_{X,\ell+1}-Z_{X,\ell}\|_2, \|Z_{S,\ell+1}-Z_{S,\ell}\|_2, \|Z_{O,\ell+1}\|_2\} \\
\le\ & \left\| \begin{bmatrix} Z_{X,\ell+1}-Z_{X,\ell} & Z_{O,\ell+1}^{\mathsf{T}} \\ Z_{O,\ell+1} & Z_{S,\ell+1}-Z_{S,\ell} \end{bmatrix} \right\|_2 \\
\le\ & \left(4\eta_\ell + 2\eta_\ell^2 \cdot \|Z_\ell\|_2 + \eta_\ell^2 \cdot \|Z_\ell\|_2 + \frac{4}{3}\eta_\ell^2 \cdot \|Z_\ell\|_2 + \frac{4}{3}\eta_\ell^3 \cdot \|Z_\ell\|_2^2 + \frac{4}{9}\eta_\ell^4 \cdot \|Z_\ell\|_2^3\right) \cdot \|Z_{O,\ell}\|_2 \\
=\ & \left(\frac{4}{9}\eta_\ell^4 \cdot \|Z_\ell\|_2^3 + \frac{4}{3}\eta_\ell^3 \cdot \|Z_\ell\|_2^2 + \frac{13}{3}\eta_\ell^2 \cdot \|Z_\ell\|_2 + 4\eta_\ell\right) \cdot \|Z_{O,\ell}\|_2^2. \qquad (84)
\end{aligned}
$$

Finally, notice that W_ℓ is skew-symmetric and $\exp(W_\ell)$ is orthogonal. Thus, the eigenvalues of Z_ℓ remain the same for all $\ell \in \mathbb{N}$, so $\|Z_\ell\|_2 = \|Z_0\|_2$ for all $\ell \in \mathbb{N}$. Therefore, replacing $\|Z_\ell\|_2$ with $\|Z_0\|_2$ in (84) gives the desirable result. $\square$

D.2 Proof of Lemma 10

A strengthened version. Here we prove a strenthened version of (43):

$$\|Z_{O,\ell}\|_2 \le \frac{1}{\|Z_0\|_2} \cdot f(\eta_0 \cdot \|Z_0\|_2) \cdot \|Z_{O,0}\|_2^{\ell+1} \qquad (85a)$$

$$\|Z_{X,\ell} - Z_{X,0}\|_2 \le \frac{1}{\|Z_0\|_2} \cdot f(\eta_0 \cdot \|Z_0\|_2) \cdot \sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2} \qquad (85b)$$

$$\|Z_{S,\ell} - Z_{S,0}\|_2 \le \frac{1}{\|Z_0\|_2} \cdot f(\eta_0 \cdot \|Z_0\|_2) \cdot \sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}. \qquad (85c)$$

More specifically, (85) implies (43) because

$$
\begin{aligned}
\frac{1}{\|Z_0\|_2} \cdot f(\eta_0 \cdot \|Z_0\|_2) &\le \frac{1}{\|Z_{X,0}\|_2} \cdot f(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + C_K)) \\
&\le \frac{1}{\|Z_{X,0}\|_2} \cdot f(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + \frac{1}{2})) = \alpha_K,
\end{aligned}
$$

following from the monotonicity of f, Lemma 18, and the definition of C_K (42): $\|Z_{O,0}\|_2 \le C_K \le \frac{1}{2}$.

Proof by induction. Now we prove (85) by induction.

1. *Base case.* When $\ell = 1$, we see from $\|Z_{O,0}\|_2 \le C_K \le \frac{3}{4\eta_0}$ and Lemma 9 that

$$
\begin{aligned}
&\max\{\|Z_{X,1} - Z_{X,0}\|_2,\ \|Z_{S,1} - Z_{S,0}\|_2,\ \|Z_{O,1}\|_2\} \\
\le\ & \frac{1}{\|Z_0\|_2}\left(\frac{4}{9}(\eta_0\|Z_0\|_2)^4 + \frac{4}{3}(\eta_0\|Z_0\|_2)^3 + \frac{13}{3}(\eta_0\|Z_0\|_2)^2 + 4(\eta_0\|Z_0\|_2)\right)\|Z_{O,0}\|_2^2 \\
\le\ & \frac{1}{\|Z_0\|_2} f(\eta_0 \cdot \|Z_0\|_2) \cdot \|Z_{O,0}\|_2^2.
\end{aligned}
$$

2. *Induction.* Suppose (85) holds for index ℓ. The proof for the induction case $\ell + 1$ takes the following three steps.

(a) First, (85b) implies that

$$
\begin{aligned}
\lambda_{\min}(Z_{X,\ell}) &\geq \lambda_{\min}(Z_{X,0}) + \lambda_{\min}(Z_{X,\ell} - Z_{X,0}) && \text{(86a)}\\
&\geq \lambda_{\min}(Z_{X,0}) - \|Z_{X,\ell} - Z_{X,0}\|_2 \\
&\geq \lambda_{\min}(Z_{X,0}) - \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2} && \text{(86b)}\\
&\geq \lambda_{\min}(Z_{X,0}) - \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \frac{\|Z_{O,0}\|_2^2}{1 - \|Z_{O,0}\|_2} \\
&\geq \lambda_{\min}(Z_{X,0}) - \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \frac{1}{\alpha_K} \cdot \frac{\frac{1}{4}\lambda_{\min}(Z_{X,0})}{1 - \frac{1}{2}} && \text{(86c)}\\
&\geq \lambda_{\min}(Z_{X,0}) - \frac{1}{2}\lambda_{\min}(Z_{X,0}) && \text{(86d)}\\
&= \frac{1}{2}\lambda_{\min}(Z_{X,0}), && \text{(86e)}
\end{aligned}
$$

where (86a) comes from Weyl's inequality, (86b) holds by the induction hypothesis, (86c) follows from the definition of C_K:

$$
\|Z_{O,0}\|_2 \leq C_K \leq \frac{1}{2}, \qquad \|Z_{O,0}\|_2 \leq C_K \leq \sqrt{\frac{1}{4\alpha_K}\lambda_{\min}(Z_{X,0})},
$$

and (86d) holds since $\alpha_K \geq \frac{1}{\|Z_0\|_2} \cdot f(\eta_0 \|Z_0\|_2)$. Similarly, we obtain

$$
\begin{aligned}
\lambda_{\min}(Z_{X,\ell}) &\leq \lambda_{\min}(Z_{X,0}) + \lambda_{\max}(Z_{X,\ell} - Z_{X,0}) \\
&\leq \lambda_{\min}(Z_{X,0}) + \|Z_{X,\ell} - Z_{X,0}\|_2 \\
&\leq \frac{3}{2}\lambda_{\min}(Z_{X,0}),
\end{aligned}
$$

and

$$
\frac{3}{2}\lambda_{\max}(Z_{S,0}) \leq \lambda_{\max}(Z_{S,0}) \leq \frac{1}{2}\lambda_{\max}(Z_{S,0}).
$$

Rearranging to get

$$
\begin{aligned}
\lambda_{\min}(Z_{X,\ell}) - \lambda_{\max}(Z_{S,\ell}) &\geq \frac{1}{2}(\lambda_{\min}(Z_{X,0}) - \lambda_{\max}(Z_{S,0})) \\
\lambda_{\min}(Z_{X,\ell}) - \lambda_{\max}(Z_{S,\ell}) &\leq \frac{3}{2}(\lambda_{\min}(Z_{X,0}) - \lambda_{\max}(Z_{S,0})),
\end{aligned}
$$

or equivalently,

$$
\frac{2}{3}\eta_0 \leq \eta_\ell \leq 2\eta_0. \tag{87}
$$

(b) Second, we show (85a) for index $\ell + 1$:

$$
\begin{aligned}
\|Z_{O,\ell+1}\|_2 &\leq \frac{f(\eta_\ell \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,\ell}\|_2^2 && \text{(88a)}\\
&\leq \frac{f(2\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,\ell}\|_2^2 && \text{(88b)}\\
&\leq \frac{f(2\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \frac{f(\eta_0 \cdot \|Z_0\|_2)^2}{\|Z_0\|_2^2} \cdot \|Z_{O,0}\|_2^{2\ell+2}, && \text{(88c)}
\end{aligned}
$$

In (88a), we use Lemma 9, the fact $Z_{X,\ell} \in \mathbb{S}^r_{++}$ and $Z_{S,\ell} \in \mathbb{S}^{n-r}_{--}$ from (a), and the fact that for any integer $\ell \geq 1$, we have

$$\eta_\ell \|Z_{O,\ell}\|_2 \leq 2\eta_0 \alpha_K \|Z_{O,0}\|_2^{\ell+1} \leq 2\eta_0 \alpha_K \|Z_{O,0}\|_2^2 \leq \frac{3}{4} \tag{89}$$

since $\|Z_{O,0}\|_2 \leq C_K \leq \min\{\frac{1}{2}, \frac{1}{\sqrt{4\eta_0\alpha_K}}\}$. Then, (88b) uses (87), and (88c) uses the induction hypothesis.

On the other hand, we see from the definition of C_K that

$$\begin{aligned}
&\|Z_{O,0}\|_2 \leq C_K \leq g^{-1}(\|Z_{X,0}\|_2^2) \\
\iff &\|Z_{O,0}\|_2 \cdot f\left(2\eta_0(\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + \|Z_{O,0}\|_2)\right) \\
&\cdot f(\eta_0(\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + \|Z_{O,0}\|_2)) \leq \|Z_{X,0}\|_2^2 && \text{(90a)} \\
\implies &\|Z_{O,0}\|_2 \cdot f\left(2\eta_0\|Z_0\|_2\right) \cdot f\left(\eta_0\|Z_0\|_2\right) \leq \|Z_{X,0}\|_2^2 && \text{(90b)} \\
\implies &\|Z_{O,0}\|_2 \cdot f(2\eta_0\|Z_0\|_2) \cdot f(\eta_0\|Z_0\|_2) \leq \|Z_0\|_2^2 \\
\iff &\|Z_{O,0}\|_2^{\ell+3} \cdot \frac{f(2\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \cdot \frac{f(\eta_0\|Z_0\|_2)^2}{\|Z_0\|_2^2} \leq \frac{f(\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,0}\|_2^{\ell+2} \\
\implies &\|Z_{O,0}\|_2^{2\ell+2} \cdot \frac{f(2\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \cdot \frac{f(\eta_0\|Z_0\|_2)^2}{\|Z_0\|_2^2} \leq \frac{f(\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,0}\|_2^{\ell+2}, && \text{(90c)}
\end{aligned}$$

where (90a) follows from the monotonicity of g, (90b) from that of f, (90c) uses the definition of C_K (so $\|Z_{O,0}\|_2 \leq C_K \leq \frac{1}{2} < 1$) and the fact $2\ell + 2 \geq \ell + 3$. Thus,

$$\|Z_{O,\ell+1}\|_2 \leq \frac{f(\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,0}\|_2^{\ell+2}.$$

(c) It remains to prove (85b) and (85c) for index $\ell + 1$. From Lemma 9 and (85a), we have

$$\begin{aligned}
&\|Z_{X,\ell+1} - Z_{X,0}\|_2 \\
\leq &\|Z_{X,\ell+1} - Z_{X,\ell}\|_2 + \|Z_{X,\ell} - Z_{X,0}\|_2 \\
\leq &\frac{f(\eta_\ell \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,\ell}\|_2^2 + \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \left(\sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}\right) \\
\leq &\frac{f(2\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,\ell}\|_2^2 + \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \left(\sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}\right) \\
\leq &\frac{f(2\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \frac{f(\eta_0\|Z_0\|_2)^2}{\|Z_0\|_2^2} \cdot \|Z_{O,0}\|_2^{2\ell+2} + \frac{f(\eta_0\|Z_0\|_2)}{\|Z_0\|_2} \left(\sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}\right) \\
\leq &\frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \|Z_{O,0}\|_2^{\ell+2} + \frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \left(\sum_{i=0}^{\ell-1} \|Z_{O,0}\|_2^{i+2}\right) \\
= &\frac{f(\eta_0 \cdot \|Z_0\|_2)}{\|Z_0\|_2} \cdot \left(\sum_{i=0}^{\ell} \|Z_{O,0}\|_2^{i+2}\right),
\end{aligned}$$

where the last inequality follows from (90c).

So we finish the proof of the strengthened inequalities (85), which implies (43). From the above mathematical induction process, we can also conclude that

$$\frac{2}{3}\eta_0 \leq \eta_\ell \leq 2\eta_0, \quad \lambda_{\min}(Z_{X,\ell}) \geq \frac{1}{2}\lambda_{\min}(Z_{X,0}) > 0 \quad \lambda_{\max}(Z_{S,\ell}) \leq \frac{1}{2}\lambda_{\max}(Z_{S,0}) < 0,$$

which follows directly from (86d) and (87).

D.3 Proof of Lemma 11

First of all, we have

$$\|W_\ell\|_2 = \|W_{O,\ell}\|_2 \overset{\text{(a)}}{\leq} \eta_\ell \|Z_{O,\ell}\|_2 \leq 2\eta_0 \|Z_{O,\ell}\|_2 \overset{\text{(b)}}{\leq} 2\eta_0 \alpha_K \|Z_{O,0}\|_2^{\ell+1} \overset{\text{(c)}}{\leq} \frac{1}{2}, \tag{91}$$

where step (a) follows from (75), step (b) holds since $Z_{O,0} \leq \min\{C_K, \frac{1}{\sqrt{4\eta_0\alpha_K}}\}$ for any integer $\ell \geq 1$, and step (c) holds for the same reason as in (88a). Then, we see that

$$\begin{aligned}
\|Q_{\ell+1} - Q_\ell\|_2 &= \|Q_\ell\left(\exp(W_\ell) - I_n\right)\|_2 \\
&\leq \|Q_\ell\|_2 \cdot \|\exp(W_\ell) - I_n\|_2 \\
&\leq \|\exp(W_\ell) - I_n\|_2 && \text{(92a)} \\
&\leq \|\exp(W_\ell) - (I_n + W_\ell)\|_2 + \|W_\ell\|_2 \\
&\leq \frac{2}{3}\|W_\ell\|_2^2 + \|W_\ell\|_2 \leq \frac{4}{3}\|W_\ell\|_2 && \text{(92b)} \\
&\leq \frac{4}{3}\eta_\ell \|Z_{O,\ell}\|_2 \leq \frac{8}{3}\eta_0 \|Z_{O,\ell}\|_2 \\
&\leq \frac{8}{3}\eta_0 \alpha_K \cdot \|Z_{O,0}\|_2^{\ell+1},
\end{aligned}$$

where (92a) holds because $Q_\ell = \prod_{i=1}^{\ell} \exp(W_i)$ is orthogonal, (92b) uses Lemma 17 and (91).

For the second inequality in the lemma, we have

$$\begin{aligned}
\|Q_\ell - (I_n + W_0)\|_2 &= \left\| Q_1 - (I_n + W_0) + \sum_{i=1}^{\ell-1} (Q_{i+1} - Q_i) \right\|_2 \\
&\leq \|Q_1 - (I_n + W_0)\|_2 + \sum_{i=1}^{\ell-1} \|Q_{i+1} - Q_i\|_2 \\
&\leq \frac{2}{3}\|W_0\|_2^2 + \frac{8}{3}\eta_0\alpha_K \sum_{i=1}^{\ell-1} \|Z_{O,0}\|_2^{i+1} && \text{(93a)} \\
&\leq \frac{2}{3}\eta_0^2 \|Z_{O,0}\|_2^2 + \frac{8}{3}\eta_0\alpha_K \sum_{i=1}^{\ell-1} \|Z_{O,0}\|_2^{i+1}, && \text{(93b)}
\end{aligned}$$

where (93a) holds because of $\eta_\ell \|Z_{O,\ell}\|_2 \leq \frac{3}{4}$ for all $\ell \in \mathbb{N}$ from (89) and (44), and (93b) follows from (75).

D.4 Proof of Lemma 12

We first prove that the limit $V_\infty := \lim_{\ell\to\infty} V_\ell$ exists. From Lemma 10, we already know that $Z_{O,\ell} \to 0$ as $\ell \to \infty$. It remains to show that $Z_{X,\ell}$ and $Z_{S,\ell}$ are also convergent. For any $\epsilon > 0$, we define

$$N := \left\lceil \frac{1}{2} \log_{\|Z_{O,0}\|_2} \left(\frac{\epsilon}{\beta}(1 - \|Z_{O,0}\|_2) \right) - 1 \right\rceil, \quad \text{where } \beta = \alpha_K^2 f(2\eta_0 \|Z_0\|_2).$$

Then, for any integers $m, n \geq N$, we have

$$
\begin{aligned}
&\|Z_{X,n} - Z_{X,m}\|_2 \\
\leq\ & \sum_{\ell=m}^{n-1} \|Z_{X,\ell+1} - Z_{X,\ell}\|_2 \\
\leq\ & \sum_{\ell=m}^{n-1} \frac{1}{\|Z_0\|_2} \cdot f(\eta_\ell \cdot \|Z_0\|_2) \cdot \|Z_{O,\ell}\|_2^2 && \text{(94a)} \\
\leq\ & \sum_{\ell=m}^{n-1} \frac{f(2\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + 0.5))}{\|Z_{X,0}\|_2} \cdot \|Z_{O,\ell}\|_2^2 && \text{(94b)} \\
\leq\ & \sum_{\ell=m}^{n-1} \frac{f(2\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + 0.5))}{\|Z_{X,0}\|_2} \cdot \alpha_K^2 \cdot \|Z_{O,0}\|_2^{2\ell+2} && \text{(94c)} \\
\leq\ & \sum_{\ell=m}^{\infty} \beta \cdot \|Z_{O,0}\|_2^{2\ell+2} \leq \beta \cdot \frac{\|Z_{O,0}\|_2^{2m+2}}{1 - \|Z_{O,0}\|_2} \leq \beta \cdot \frac{\|Z_{O,0}\|_2^{2N+2}}{1 - \|Z_{O,0}\|_2} \leq \epsilon, && \text{(94d)}
\end{aligned}
$$

Here, (94a) holds because of Lemma 9 and the fact that $\eta_\ell \cdot \|Z_{O,\ell}\|_2 \leq \frac{3}{4}$, for any integer $\ell \geq 1$ (see (89)), (94b) uses Lemma 18, the monotonicity of f, and the fact $\|Z_{O,0}\|_2 \leq C_K \leq \frac{1}{2}$, (94c) uses (43a), and (94d) follows directly from the definition of β and N.

Thus, $\{Z_{X,\ell}\}_{\ell=1}^{\infty} \subseteq \mathbb{S}^r$ is a Cauchy sequence and the limit $Z_{X,\infty} := \lim_{\ell\to\infty} Z_{X,\ell}$ exists. The same argument applies to $Z_{S,\infty}$. Similarly, from Lemma 11, we conclude that $\{Q_\ell\}_{\ell=1}^{\infty}$ is convergent to Q_∞, and thus the limit

$$
V_\infty := \lim_{\ell\to\infty} Q_\ell \begin{bmatrix} Z_{X,\ell} & 0 \\ 0 & 0 \end{bmatrix} Q_\ell^\top = Q_\infty \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0 \end{bmatrix} Q_\infty^\top
$$

exists.

It only remains to show that $V_\infty = \Pi_{\mathbb{S}_+^n}(Z+H)$. It readily follows from

$$
\begin{aligned}
Z_\ell &= \exp(W_\ell)^\top \begin{bmatrix} Z_{X,\ell} & Z_{O,\ell}^\top \\ Z_{O,\ell} & Z_{S,\ell} \end{bmatrix} \exp(W_\ell) \\
&= (\exp(W_{\ell-1})\exp(W_\ell))^\top \begin{bmatrix} Z_{X,\ell-1} & Z_{O,\ell-1}^\top \\ Z_{O,\ell-1} & Z_{S,\ell-1} \end{bmatrix} \exp(W_{\ell-1})\exp(W_\ell) \\
&\ \ \vdots \\
&= \left(\prod_{i=0}^{\ell} \exp(W_i)\right)^\top \begin{bmatrix} Z_{X,0} & Z_{O,0}^\top \\ Z_{O,0} & Z_{S,0} \end{bmatrix} \left(\prod_{i=0}^{\ell} \exp(W_i)\right) \\
&= Q_\ell^\top (Z+H) Q_\ell.
\end{aligned}
$$

Thus, we conclude that

$$
Z_\infty = \lim_{\ell\to\infty} Z_\ell = Q_\infty^\top (Z+H) Q_\infty.
$$

Recall that for any $\ell \in \mathbb{N}$, $\exp(W_\ell)$ is orthogonal, so $Q_\infty = \lim_{\ell\to\infty} \prod_{i=0}^{\ell} \exp(W_i)$ is also orthogonal. It implies that

$$
\Pi_{\mathbb{S}_+^n}(Z+H) = Q_\infty \cdot \Pi_{\mathbb{S}_+^n}(Z_\infty) \cdot Q_\infty^\top.
$$

On the other hand, recall from Lemma 10 that $\lambda_{\min}(Z_{X,\ell}) \geq \frac{1}{2}\lambda_{\min}(Z_{X,0}) > 0$ and $-\lambda_{\max}(Z_{S,\ell}) \geq$

$-\frac{1}{2}\lambda_{\max}(Z_{S,0}) > 0$, so we have $Z_{X,\infty} \in \mathbb{S}^r_{++}$, $-Z_{S,\infty} \in \mathbb{S}^{n-r}_{--}$, and

$$\begin{aligned}\Pi_{\mathbb{S}^n_+}(Z_\infty) &= \Pi_{\mathbb{S}^n_{++}}\left(\begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & Z_{S,\infty}\end{bmatrix}\right) = \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0\end{bmatrix}, \\ \Pi_{\mathbb{S}^n_+}(Z+H) &= Q_\infty \begin{bmatrix} Z_{X,\infty} & 0 \\ 0 & 0\end{bmatrix} Q_\infty^\mathsf{T} = V_\infty.\end{aligned}$$

This concludes the proof.

D.5 Proof of Lemma 13

We prove the four conclusions one-by-one.

1. Since $\|H\|_2 \leq \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\}$, we have

$$\lambda_{\min}(Z_{X,0}) \geq \lambda_{\min}(\Lambda_X) - \|Z_X\|_2 \geq \lambda_r - \|H\|_2 \geq \frac{1}{2}\lambda_r > 0$$

by Weyl's inequality and Lemma 18. Symmetrically, we obtain $\lambda_{\max}(Z_{S,0}) \leq \frac{1}{2}\lambda_{r+1} < 0$. Similarly, we can obtain an upper bound for $\lambda_{\min}(Z_{X,0})$ and a lower bound for $\lambda_{\max}(Z_{S,0})$:

$$\lambda_{\min}(Z_{X,0}) \leq \frac{3}{2}\lambda_r, \qquad \lambda_{\max}(Z_{S,0}) \leq \frac{3}{2}\lambda_{r+1}.$$

2. Now we bound η_0. On one hand, we have

$$\eta_0 = \frac{d}{\lambda_{\min}(Z_{X,0}) - \lambda_{\max}(Z_{S,0})} \leq \frac{d}{\frac{1}{2}\lambda_r - \frac{1}{2}\lambda_{r+1}} = \frac{2d}{\lambda_r - \lambda_{r+1}} =: \eta_{0,f}. \tag{95}$$

On the other hand, we have an lower bound for η_0:

$$\eta_0 \geq \frac{d}{1.5\lambda_r - 1.5\lambda_{r+1}} = \frac{2d}{3(\lambda_r - \lambda_{r+1})}.$$

3. To bound α_K, we first notice from $\|H\|_2 \leq \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\}$ that

$$\begin{aligned}\|Z+H\|_2 &\geq \|Z\|_2 - \|H\|_2 \geq \|Z\|_2 - 0.5\min\{\lambda_r, -\lambda_{r+1}\}, \\ \|Z+H\|_2 &\leq \|Z\|_2 + \|H\|_2 \leq \|Z\|_2 + 0.5\min\{\lambda_r, -\lambda_{r+1}\}.\end{aligned}$$

Then, from the definition of α_K (42), we deduce that

$$\begin{aligned}\alpha_K &= \frac{f\left(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + 0.5)\right)}{\|Z_{X,0}\|_2} \\ &\leq \frac{f\left(\frac{2d}{3(\lambda_r - \lambda_{r+1})} \cdot (2(\lambda_1 - \lambda_n) + \min\{\lambda_r, -\lambda_{r+1}\} + 0.5)\right)}{\lambda_1 - 0.5\lambda_r} \\ &=: \alpha_{K,f},\end{aligned} \tag{96}$$

where the inequality uses Weyl's inequality:

$$\begin{aligned}&\|Z_0\|_2 \geq \max\{\|Z_{X,0}\|_2, \|Z_{S,0}\|_2\}, \\ &\|Z_0\|_2 \leq \|Z\|_2 + \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\} \leq (\lambda_1 - \lambda_n) + \frac{1}{2}\min\{\lambda_r, -\lambda_{r+1}\}\end{aligned}$$

and recall $\|Z_{X,0}\|_2 \geq \lambda_1 - \frac{1}{2}\lambda_r$.

4. It remains to find a lower bound for C_K. Since C_K is the minimum of five terms, we bound each of them one-by-one.

   - $C_1 = \frac{1}{2}$, which is trivial.
   - $C_2 = g^{-1}(\|Z_{X,0}\|_2^2)$. We see from the definition of g that

$$
\begin{aligned}
g(y) &= y \cdot f\left(2\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + y)\right) \cdot f\left(\eta_0 \cdot (\|Z_{X,0}\|_2 + \|Z_{S,0}\|_2 + y)\right) \\
&\leq y \cdot f\left(2\eta_{0,f} \cdot (2(\lambda_1 - \lambda_n) + \min\{\lambda_r, -\lambda_{r+1}\} + y)\right) \\
&\quad \cdot f\left(\eta_{0,f} \cdot (2(\lambda_1 - \lambda_n) + \min\{\lambda_r, -\lambda_{r+1}\} + y)\right) =: g_1(y).
\end{aligned}
$$

     It is clear that g_1 is monotonically increasing on $[0, \infty)$ and $g_1(0) = 0$. Combined with the fact that $\|Z_{X,0}\|_2 \geq \lambda_1 - \frac{1}{2}\lambda_r$, we conclude that

$$
g^{-1}(\|Z_{X,0}\|_2^2) \geq g_1^{-1}(\|Z_{X,0}\|_2^2) \geq g_1^{-1}((\lambda_1 - 0.5\lambda_r)^2) =: C_{2,f} > 0,
$$

     which serves as a lower bound for C_2.
   - $C_3 = \frac{3}{8\eta_0} \geq \frac{3}{8\eta_{0,f}} =: C_{3,f}$.
   - $C_4 = \frac{1}{\sqrt{4\eta_0 \alpha_K}} \geq \frac{1}{\sqrt{4\eta_{0,f}\alpha_{K,f}}} =: C_{4,f}$.
   - Lastly, we have

$$
\begin{aligned}
C_5 &= \sqrt{\frac{1}{4\alpha_K} \min\{\lambda_{\min}(Z_{X,0}), -\lambda_{\max}(Z_{S,0})\}} \\
&\geq \sqrt{\frac{1}{4\alpha_{K,f}} \min\{\frac{1}{2}\lambda_{\min}(Z_{X,0}), -\frac{1}{2}\lambda_{\max}(Z_{S,0})\}} =: C_{5,f}.
\end{aligned}
$$

   Therefore, $C_{K,f} := \min\{\frac{1}{2}, C_{2,f}, C_{3,f}, C_{4,f}, C_{5,f}\}$ serves as a lower bound for C_K.

D.6 Proof of Lemma 14

We first show $\Theta_0 \circ H_O$ is the solution of

$$
W_O \Lambda_X - \Lambda_S W_O = H_O.
$$

To see this, expand both sides of the equity and consider the (i,j)th element, $i \in \{r+1, \ldots, n\}$ and $j \in \{1, \ldots, r\}$:

$$
W_{O,i,j} \cdot \lambda_j - \lambda_i \cdot W_{O,i,j} = H_{O,i,j} \implies W_{O,i,j} = \frac{1}{\lambda_j - \lambda_i} H_{O,i,j} \implies W_O = \Theta_0 \circ H_O.
$$

Second, we show that the perturbations H_X and H_S only affect the second- and higher-order terms. To see this, we explicitly write down $W_{O,0}$ as:

$$
\begin{aligned}
& W_{O,0}(\Lambda_X + H_X) - (\Lambda_S + H_S) W_{O,0} = H_O \\
\implies \quad & \mathrm{vec}(W_{O,0}) = (A + \Delta A)^{-1} \mathrm{vec}(H_O),
\end{aligned}
$$

where $A := I_r \otimes (-\Lambda_S) + \Lambda_X \otimes I_{n-r}$ and $\Delta A := I_r \otimes (-H_S) + H_X \otimes I_{n-r}$. We bound $\|\Delta A\|_2$ as

$$
\begin{aligned}
\|\Delta A\|_2 \leq \|\Delta A\|_\mathsf{F} &= \|I_r \otimes (-H_S) + H_X \otimes I_{n-r}\|_\mathsf{F} \\
&\leq \|I_r \otimes (-H_S)\|_\mathsf{F} + \|H_X \otimes I_{n-r}\|_\mathsf{F} \\
&= \|I_r\|_\mathsf{F} \cdot \|H_S\|_\mathsf{F} + \|I_{n-r}\|_\mathsf{F} \cdot \|H_X\|_\mathsf{F} \\
&= r\|H_S\|_\mathsf{F} + (n-r)\|H_X\|_\mathsf{F} \\
&\leq n(\|H_X\|_\mathsf{F} + \|H_S\|_\mathsf{F}), \qquad (97)
\end{aligned}
$$

and bound $\|A^{-1}\|_2$ as

$$\|A^{-1}\|_2 = \left(\lambda_{\min}\left(I_r \otimes (-\Lambda_S) + \Lambda_X \otimes I_{n-r}\right)\right)^{-1} = \frac{1}{\lambda_r - \lambda_{r+1}}, \tag{98}$$

which follows from [52, Theorem 2.5] and the fact that $A \in \mathbb{S}^n_{++}$. Combining (97) and (98) gives

$$\|A^{-1}\Delta A\|_2 \leq \frac{n}{\lambda_r - \lambda_{r+1}}(\|H_X\|_\mathsf{F} + \|H_S\|_\mathsf{F}) \leq \frac{nd}{\lambda_r - \lambda_{r+1}}(\|H_X\|_2 + \|H_S\|_2) \leq \frac{1}{2} \tag{99}$$

where the last inequliaty is from the assumption of the lemma. Therefore, we have

$$\begin{aligned}
&\|W_{O,0} - \Theta_0 \circ H_O\|_2 \\
\leq\ &\|W_{O,0} - \Theta_0 \circ H_O\|_\mathsf{F} \\
=\ &\|\mathrm{vec}(W_{O,0}) - \mathrm{vec}(\Theta_0 \circ H_O)\|_2 \\
=\ &\|(A + \Delta A)^{-1}\mathrm{vec}(H_O) - A^{-1}\mathrm{vec}(H_O)\|_2 \\
\leq\ &\|A^{-1}\mathrm{vec}(H_O)\|_2 \cdot \|(I + A^{-1}\Delta A)^{-1} - I\|_2 \\
=\ &\|A^{-1}\mathrm{vec}(H_O)\|_2 \cdot \left\|\sum_{i=0}^{\infty}(-A^{-1}\Delta A)^i - I\right\|_2 && \text{(100a)} \\
\leq\ &\|A^{-1}\mathrm{vec}(H_O)\|_2 \cdot \sum_{i=1}^{\infty}\|A^{-1}\Delta A\|_2^i \\
\leq\ &\|A^{-1}\|_2 \cdot \|\mathrm{vec}(H_O)\|_2 \cdot \frac{\|A^{-1}\Delta A\|_2}{1 - \|A^{-1}\Delta A\|_2} \\
\leq\ &\frac{1}{\lambda_r - \lambda_{r+1}} \cdot \|H_O\|_2 \cdot 2 \cdot \frac{1}{\lambda_r - \lambda_{r+1}} \cdot \|\Delta A\|_2 && \text{(100b)} \\
\leq\ &\frac{2nd}{(\lambda_r - \lambda_{r+1})^2} \cdot \|H_O\|_2 \cdot (\|H_X\|_2 + \|H_S\|_2). && \text{(100c)}
\end{aligned}$$

In (100a), the expansion of the Neumann series is valid because $\|A^{-1}\Delta A\|_2 < \frac{1}{2}$ (see (98)); (100b) uses (98) and (99); and finally (100c) uses (97).

Appendix E Additional Numerical Results

Figure 14 presents additional numerical results for the Hamming set problems [48]. Figure 15 reports additional examples for the BQP problems, with $c \sim \mathcal{N}(0, I_n)$ and random (standard Gaussian) initial guess, and Figure 16 reports for the same problems as in Figure 15, but with all-zeros initialization. Last, Figure 17 shows numerical results for BQP problems with $c = 0$ and random (standard Gaussian) initial guess.

References

[1] Farid Alizadeh, Jean-Pierre A Haeberly, and Michael L Overton. Complementarity and nondegeneracy in semidefinite programming. *Mathematical Programming*, 77(1):111–128, 1997.

[2] Farid Alizadeh, Jean-Pierre A Haeberly, and Michael L Overton. Primal–dual interior-point methods for semidefinite programming: Convergence rates, stability and numerical results. *SIAM Journal on Optimization*, 8(3):746–768, 1998.

[3] 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):133–184, 2023.

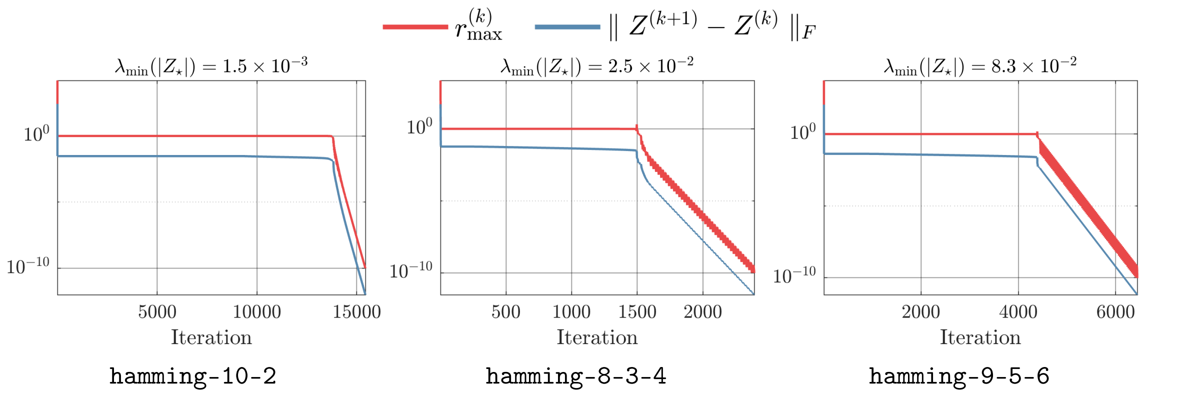


Figure 14: More Hamming graph problems with with random (standard Gaussian) initial guess. In all cases, the converging $Z_\star$ is nonsingular.

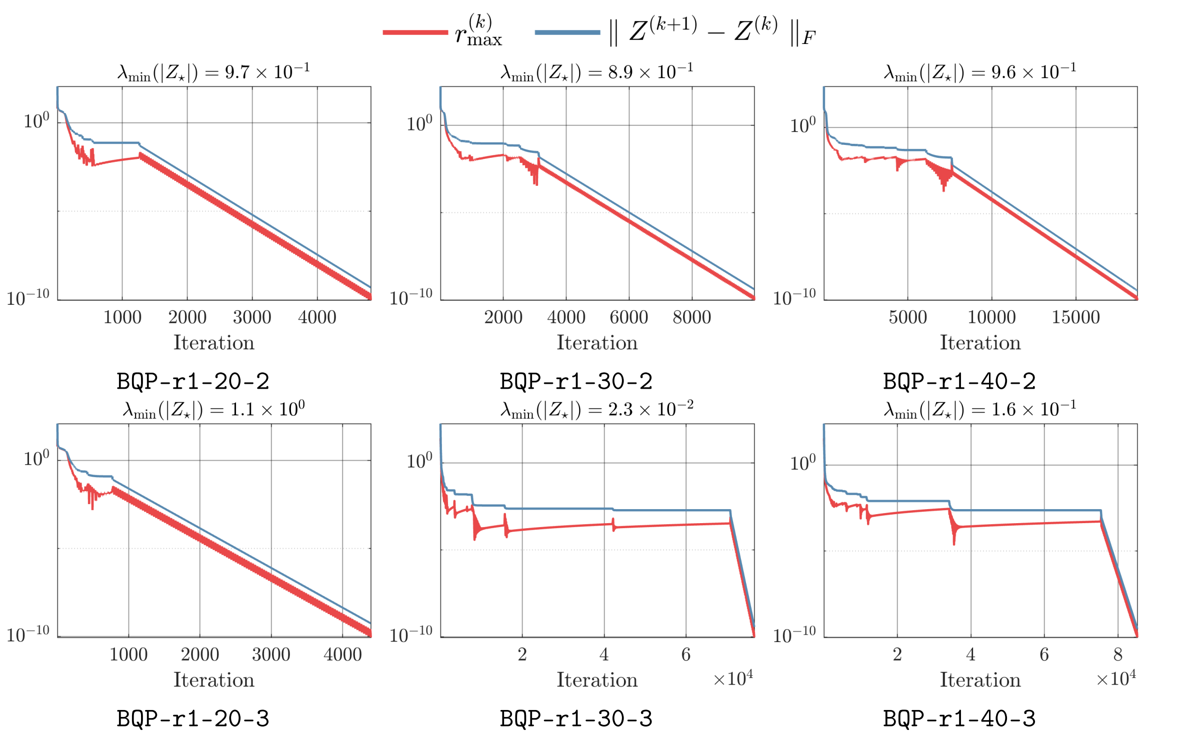


Figure 15: Additional Random BQP problems with $c \sim \mathcal{N}(0, I_n)$ and random (standard Gaussian) initialization. In all cases, the converging $Z_\star$ is nonsingular.

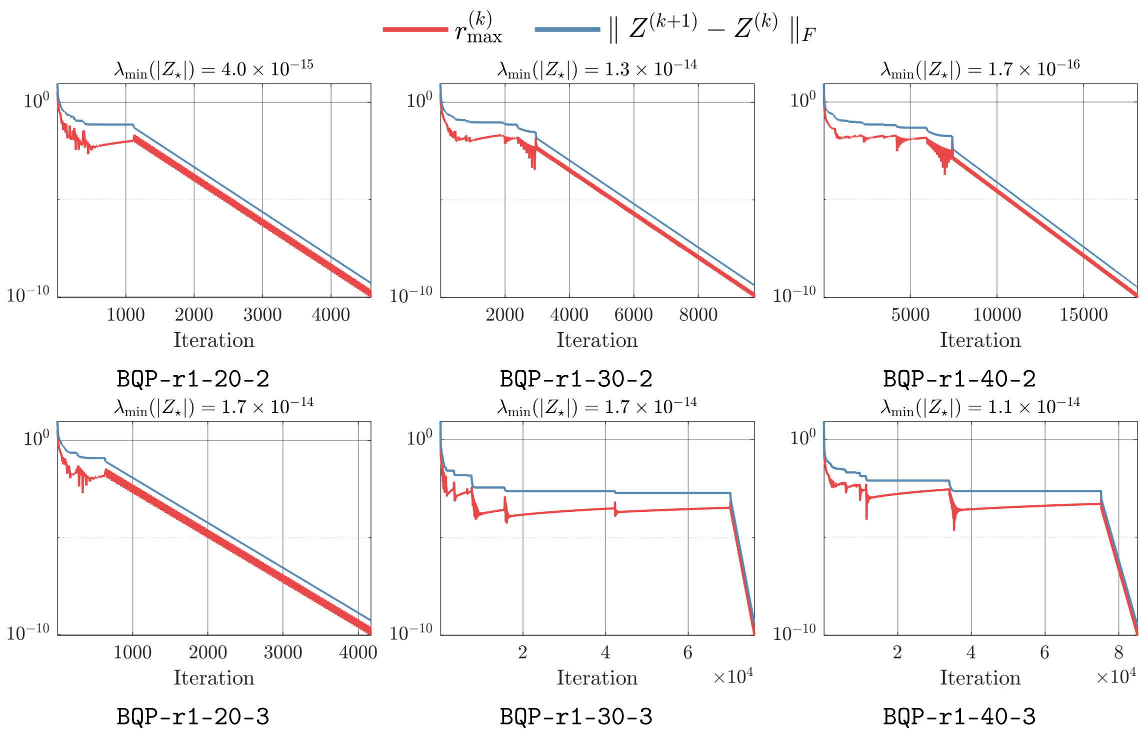


Figure 16: Additional Random BQP problems with $c \sim \mathcal{N}(0, I_n)$ and all-zeros initialization. In all cases, the converging $Z_\star$ is singular.

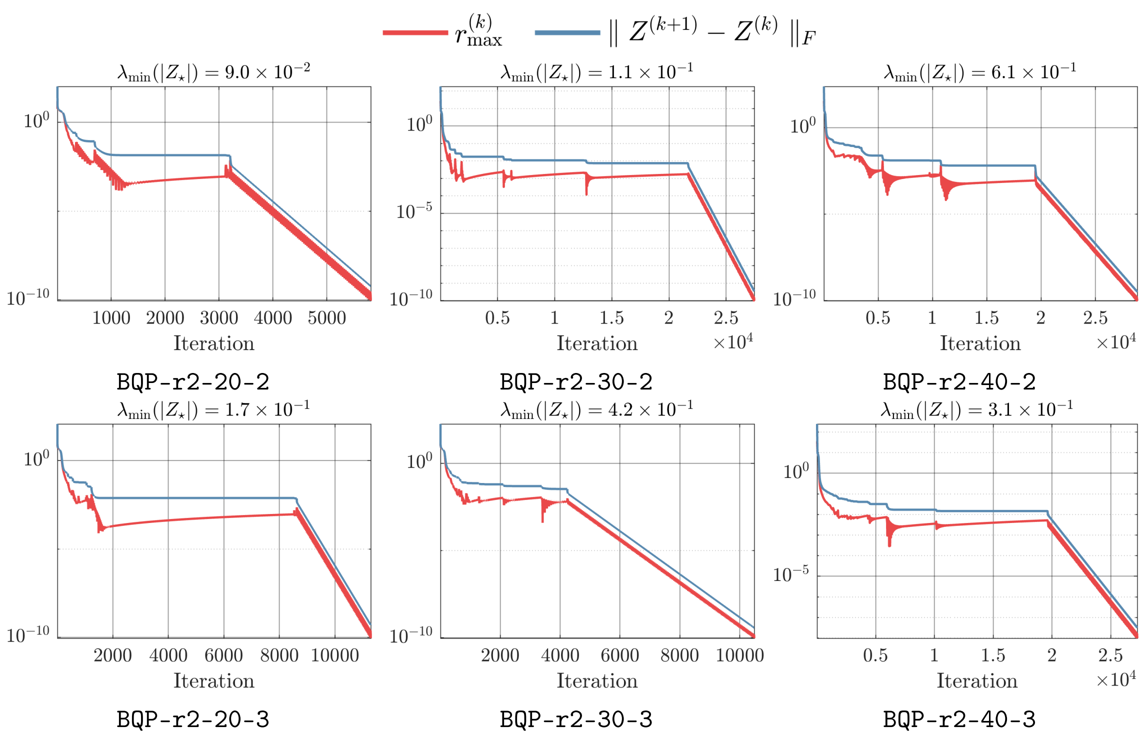


Figure 17: Additional random BQP problems with $c = 0$ and random (standard Gaussian) initialization. In all cases, the converging $Z_\star$ is nonsingular.

[4] Mosek ApS. MOSEK optimization toolbox for MATLAB. *User's Guide and Reference Manual, Version*, 4(1), 2019.

[5] Heinz H Bauschke, JY Bello Cruz, Tran TA Nghia, Hung M Pha, and Xianfu Wang. Optimal rates of linear convergence of relaxed alternating projections and generalized Douglas–Rachford methods for two subspaces. *Numerical Algorithms*, 73:33–76, 2016.

[6] Heinz H Bauschke and Patrick L Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, Berlin, Heidelberg, 2017.

[7] Nicolas Boumal, Vlad Voroninski, and Afonso Bandeira. The non-convex Burer–Monteiro approach works on smooth semidefinite programs. *Advances in Neural Information Processing Systems*, 29, 2016.

[8] Samuel Burer and Renato DC Monteiro. A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Mathematical Programming*, 95(2):329–357, 2003.

[9] Zi Xian Chan and Defeng Sun. Constraint nondegeneracy, strong regularity, and nonsingularity in semidefinite programming. *SIAM Journal on Optimization*, 19(1):370–396, 2008.

[10] Liang Chen, Xudong Li, Defeng Sun, and Kim-Chuan Toh. On the equivalence of inexact proximal ALM and ADMM for a class of convex composite programming. *Mathematical Programming*, 185(1):111–161, 2021.

[11] Liang Chen, Defeng Sun, and Kim-Chuan Toh. An efficient inexact symmetric Gauss–Seidel based majorized ADMM for high-dimensional convex composite conic programming. *Mathematical Programming*, 161:237–270, 2017.

[12] Ying Cui, Defeng Sun, and Kim-Chuan Toh. On the asymptotic superlinear convergence of the augmented Lagrangian method for semidefinite programming with multiple solutions. *arXiv preprint arXiv:1610.00875*, 2016.

[13] Damek Davis and Wotao Yin. Faster convergence rates of relaxed Peaceman-Rachford and ADMM under regularity assumptions. *Mathematics of Operations Research*, 42(3):783–805, August 2017.

[14] Timothy A. Davis and Yifan Hu. The University of Florida sparse matrix collection. *ACM Transactions on Mathematical Software (TOMS)*, 38(1):1–25, 2011.

[15] 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, March 2016.

[16] Lijun Ding and Benjamin Grimmer. Revisiting spectral bundle methods: Primal–dual (sub) linear convergence rates. *SIAM Journal on Optimization*, 33(2):1305–1332, 2023.

[17] Lijun Ding and Madeleine Udell. A strict complementarity approach to error bound and sensitivity of solution of conic programs. *Optimization Letters*, 17(7):1551–1574, 2023.

[18] Dmitriy Drusvyatskiy and Adrian S. Lewis. Optimality, identifiability, and sensitivity. *Mathematical Programming*, 147(1):467–498, 2014.

[19] Giovanni Fantuzzi and Federico Fuentes. Global minimization of polynomial integral functionals. *SIAM Journal on Scientific Computing*, 46(4):A2123–A2149, 2024.

[20] Michael Garstka, Mark Cannon, and Paul Goulart. Cosmo: A conic operator splitting method for convex conic problems. *Journal of Optimization Theory and Applications*, 190(3):779–810, 2021.

[21] Shai Gertler, Zeyu Kuang, Colin Christie, Hao Li, and Owen D Miller. Many photonic design problems are sparse QCQPs. *Science Advances*, 11(1):eadl3237, 2025.

[22] 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.

[23] Haoyu Han and Heng Yang. Building rome with convex optimization. *arXiv preprint arXiv:2502.04640*, 2025.

[24] 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. *arXiv preprint arXiv:2403.09133*, 2024.

[25] Christoph Helmberg and Franz Rendl. A spectral bundle method for semidefinite programming. *SIAM Journal on Optimization*, 10(3):673–696, 2000.

[26] Alan J. Hoffman. On approximate solutions of systems of linear inequalities. In *Selected Papers Of Alan J Hoffman: With Commentary*, pages 174–176. World Scientific, Singapore, 2003.

[27] Mingyi Hong and Zhi-Quan Luo. On the linear convergence of the alternating direction method of multipliers. *Mathematical Programming*, 162(1):165–199, 2017.

[28] Lei Huang, Shucheng Kang, Jie Wang, and Heng Yang. Sparse polynomial optimization with unbounded sets. *arXiv preprint arXiv:2401.15837*, 2024.

[29] Xin Jiang and Lieven Vandenberghe. Bregman primal–dual first-order method and application to sparse semidefinite programming. *Computational Optimization and Applications*, 81(1):127–159, 2022.

[30] Shucheng Kang, Guorui Liu, and Heng Yang. Global contact-rich planning with sparsity-rich semidefinite relaxations. *arXiv preprint arXiv:2502.02829*, 2025.

[31] Shucheng Kang, Xiaoyang Xu, Jay Sarva, Ling Liang, and Heng Yang. Fast and certifiable trajectory optimization. In *International Workshop on the Algorithmic Foundations of Robotics*, 2024.

[32] Jean B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[33] Adrian S. Lewis. Active sets, nonsmoothness, and sensitivity. *SIAM Journal on Optimization*, 13(3):702–725, 2002.

[34] Yongfeng Li, Zaiwen Wen, Chao Yang, and Ya-Xiang Yuan. A semismooth newton method for semidefinite programs and its applications in electronic structure calculations. *SIAM Journal on Scientific Computing*, 40(6):A4131–A4157, 2018.

[35] 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*, 172:874–913, 2017.

[36] Feng-Yi Liao, Lijun Ding, and Yang Zheng. An overview and comparison of spectral bundle methods for primal and dual semidefinite programs. *arXiv preprint arXiv:2307.07651*, 2023.

[37] Feng-Yi Liao, Lijun Ding, and Yang Zheng. Inexact augmented Lagrangian methods for conic optimization: Quadratic growth and linear convergence. *Advances in Neural Information Processing Systems*, 37:41013–41050, 2024.

[38] Feng-Yi Liao and Yang Zheng. A bundle-based augmented Lagrangian framework: Algorithm, convergence, and primal–dual principles. *arXiv preprint arXiv:2502.08835*, 2025.

[39] Yongchao Liu, Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. Partial error bound conditions and the linear convergence rate of the alternating direction method of multipliers. *SIAM Journal on Numerical Analysis*, 56(4):2095–2123, January 2018.

[40] Haihao Lu and Jinwen Yang. On the geometry and refined rate of primal–dual hybrid gradient for linear programming. *Mathematical Programming*, pages 1–39, 2024.

[41] Victor Magron and Jie Wang. *Sparse Polynomial Optimization: Theory and Practice*. World Scientific, Singapore, 2023.

[42] Hans D. Mittelmann. Several SDP-codes on sparse and other SDP problems, 2006.

[43] Jiawang Nie. *Moment and Polynomial Optimization*. SIAM, Philadelphia, PA, 2023.

[44] Robert Nishihara, Laurent Lessard, Ben Recht, Andrew Packard, and Michael Jordan. A general analysis of the convergence of ADMM. In *International Conference on Machine learning*, pages 343–352. PMLR, 2015.

[45] Daniel O'Connor and Lieven Vandenberghe. On the equivalence of the primal–dual hybrid gradient method and Douglas–Rachford splitting. *Mathematical Programming*, 179(1-2):85–108, 2020.

[46] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. SCS: Splitting conic solver, version 3.2.4. https://github.com/cvxgrp/scs, November 2023.

[47] Beresford N. Parlett. *The symmetric eigenvalue problem*. SIAM, Philadelphia, PA, 1998.

[48] Gabor Pataki and Stefan Schmieta. The DIMACS library of semidefinite-quadratic-linear programs. 1999.

[49] Michael J. D. Powell. A method for nonlinear constraints in minimization problems. *Optimization*, pages 283–298, 1969.

[50] R Tyrrell Rockafellar. Augmented lagrangians and applications of the proximal point algorithm in convex programming. *Mathematics of operations research*, 1(2):97–116, 1976.

[51] Ernest K Ryu and Wotao Yin. *Large-scale convex optimization: algorithms & analyses via monotone operators*. Cambridge University Press, Cambridge, 2022.

[52] Kathrin Schacke. On the Kronecker product. Master's thesis, University of Waterloo, 2004.

[53] Jos F. Sturm. Error bounds for linear matrix inequalities. *SIAM Journal on Optimization*, 10(4):1228–1248, 2000.

[54] Defeng Sun and Jie Sun. Semismooth matrix-valued functions. *Mathematics of Operations Research*, 27(1):150–169, 2002.

[55] Tianyun Tang and Kim-Chuan Toh. A feasible method for general convex low-rank SDP problems. *arXiv preprint arXiv:2312.07908*, 2023.

[56] Jie Wang and Liangbing Hu. Solving low-rank semidefinite programs via manifold optimization. *arXiv preprint arXiv:2303.01722*, 2023.

[57] Jie Wang, Victor Magron, and Jean B Lasserre. Certifying global optimality of ac-opf solutions via sparse polynomial optimization. *Electric Power Systems Research*, 213:108683, 2022.

[58] Zaiwen Wen, Donald Goldfarb, and Wotao Yin. Alternating direction augmented Lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3):203–230, 2010.

[59] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, Berlin, Heidelberg, 2012.

[60] Stephen J. Wright. Identifiable surfaces in constrained optimization. *SIAM Journal on Control and Optimization*, 31(4):1063–1079, 1993.

[61] Heng Yang and Luca Carlone. A quaternion-based certifiably optimal solution to the wahba problem with outliers. In *International Conference on Computer Vision*, pages 1665–1674, 2019.

[62] Heng Yang and Luca Carlone. Certifiably optimal outlier-robust geometric perception: Semidefinite relaxations and scalable global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(3):2816–2834, 2022.

[63] 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):409–472, 2023.

[64] 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.

[65] Xiaoming Yuan, Shangzhi Zeng, and Jin Zhang. Discerning the linear convergence of ADMM for structured convex optimization through the lens of variational analysis. *Journal of Machine Learning Research*, 21(83):1–75, 2020.

[66] Moslem Zamani, Hadi Abbaszadehpeivasti, and Etienne de Klerk. The exact worst-case convergence rate of the alternating direction method of multipliers. *Mathematical Programming*, 208(1):243–276, 2024.

[67] Ning Zhang, Jia Wu, and Liwei Zhang. A linearly convergent majorized ADMM with indefinite proximal terms for convex composite programming and its applications. *Mathematics of Computation*, 89(324):1867–1894, 2020.

[68] Xin-Yuan 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.

[69] Yang Zheng, Giovanni Fantuzzi, Antonis Papachristodoulou, Paul Goulart, and Andrew Wynn. Fast ADMM for homogeneous self-dual embedding of sparse SDPs. *IFAC-PapersOnLine*, 50(1):8411–8416, 2017.