Noname manuscript No.
(will be inserted by the editor)

An Inexact Augmented Lagrangian Method for Semidefinite Complementarity Optimization

Wanqi Hou · Yanzhen Peng · Chao Ding

This version: November 10, 2026

Abstract Semidefinite complementarity constraints arise in applications such as inverse semidefinite programming and rank-constrained matrix optimization. We develop a structure-preserving augmented Lagrangian method (SPALM) for mathematical programs with these constraints. The method augments the smooth constraints and retains matrix complementarity in its subproblems. We handle these subproblems by an inexact proximal alternating framework with verifiable descent and stationarity conditions. With vanishing absolute subproblem errors, feasible accumulation points of infinite outer sequences are asymptotically M-stationary. Under AM-regularity, these points are M-stationary; we prove that active-set SDCMPCC-LICQ is sufficient for this regularity. For each AL subproblem, the full stationarity residual converges to zero along infinite accepted inner sequences if the initial sublevel set is bounded, the proximal parameters are bounded above and away from zero, and the block errors vanish. If the block errors are also square-summable, an $O(T^{-1/2})$ best-residual bound holds over T accepted iterations. Huber–Stiefel and fixed-basis procedures generate feasible matrix candidates, whose acceptance is verified by descent tests and computable limiting-normal certificates. For rank-constrained correlation matrix approximation, we construct a lifting that preserves the projected feasible set and global optimal value and satisfies active-set SDCMPCC-LICQ at every feasible point. The stationarity guarantees apply to this lifted model. Finite-budget experiments on this lifting and a single-equality observation-fitting inverse SDP meet the prescribed numerical feasibility and stationarity tests in all 120 runs.

Key words: semidefinite cone complementarity constraints; augmented Lagrangian method; asymptotic stationarity; proximal alternating method

1 Introduction

Mathematical programs with equilibrium constraints (MPECs) arise in equilibrium models, variational inequalities, and bilevel programming, with applications in engineering and economics. Mathematical programs with complementarity constraints (MPCCs) form a basic subclass [27,33]. Complementarity constraints also enter optimality-based reformulations of constrained bilevel problems. Bilevel models appear, for example, in hyperparameter optimization and meta-learning [12], and in domain reweighting for large language model pretraining [11]. These connections motivate the study of complementarity-constrained optimization. Although optimality theory and numerical methods for vector MPCCs are well developed, optimization with matrix-valued complementarity constraints has received less attention.

Matrix complementarity arises naturally in inverse optimization, where observed decisions are used to infer parameters of a forward optimization model [8]. For a linear semidefinite forward problem, the Karush–Kuhn–

Wanqi Hou
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences, Beijing, China
E-mail: houwanqi19@mails.ucas.ac.cn

Yanzhen Peng
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
School of Advanced Interdisciplinary Sciences, University of Chinese Academy of Sciences, Beijing, China
E-mail: pengyanzhen23@mails.ucas.ac.cn

Chao Ding
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China
E-mail: dingchao@amss.ac.cn

Tucker (KKT) conditions consist of primal–dual feasibility and complementarity between a positive semidefinite primal matrix X and a positive semidefinite dual slack S:

$$X \succeq 0, \qquad S \succeq 0, \qquad XS = 0.$$

Imposing these KKT conditions in an inverse model introduces a semidefinite complementarity constraint; see, for example, [37].

Low-rank correlation matrix approximation provides another application. In financial model calibration, one seeks a correlation matrix that approximates a given target under a prescribed rank bound [38,7,23]. The approximating matrix must be positive semidefinite and have unit diagonal. An auxiliary matrix allows the rank restriction to be expressed through semidefinite complementarity and additional constraints [9]; Section 6.2 develops the lifting used here. Further applications, including certain robust optimization models, can be found in [9].

Motivated by these applications, we develop a numerical method for mathematical programs with semidefinite cone complementarity constraints (SDCMPCCs) of the form

$$\min_{w} f(w) \quad \text{s.t.} \quad g(w) \le 0,\ h(w) = 0,\ \mathbb{S}^n_+ \ni G(w) \perp H(w) \in \mathbb{S}^n_-, \tag{1}$$

Here $w \in \mathcal{Z}$, where $\mathcal{Z}$ is a finite-dimensional real Euclidean space. The mappings $f : \mathcal{Z} \to \mathbb{R}$, $g : \mathcal{Z} \to \mathbb{R}^p$, $h : \mathcal{Z} \to \mathbb{R}^q$, and $G, H : \mathcal{Z} \to \mathbb{S}^n$ are continuously differentiable. SDCMPCCs are matrix analogues of MPCCs and include them as a special case. The complementarity constraint is described by the set

$$\Omega := \operatorname{gph} N_{\mathbb{S}^n_+} = \big\{(Y, Z) \in \mathbb{S}^n_+ \times \mathbb{S}^n_- :\ YZ = 0\big\}.$$

Thus the complementarity condition in (1) is $(G(w), H(w)) \in \Omega$. Even for vector MPCCs, complementarity causes standard constraint qualifications to fail, and classical KKT conditions need not hold at local minimizers [27,33]. SDCMPCCs combine this degeneracy with the nonpolyhedral geometry of $\mathbb{S}^n_+$ and the nonconvexity of Ω. Stationarity analysis and algorithm design therefore require the geometry of the semidefinite complementarity set.

On the theoretical side, Ding et al. [9] showed that the Robinson constraint qualification fails in the direct conic formulation of SDCMPCCs. Their proximal and limiting normal-cone formulas for Ω give explicit spectral characterizations of S- and M-stationarity, while a projection reformulation yields C-stationarity conditions. Wu et al. [36] established a tangent-cone formula and characterized B-stationarity. They also studied its relationships with S-, M-, and C-stationarity under SDCMPCC-tailored constraint qualifications, together with related second-order sufficient conditions. Madariaga and Ramírez [28] study stationarity for general cone complementarity constraints from a facial viewpoint. These results provide the geometric basis for the stationarity conditions used in SDCMPCC algorithms.

Algorithmic research on vector MPCCs includes regularization and relaxation methods [34,21,22], penalty and exact-penalty methods [30,16,25], and augmented Lagrangian methods [19,3,15]. In particular, Guo and Deng [15] retain vector complementarity explicitly in the augmented Lagrangian subproblems and solve them by a nonmonotone projected gradient method. This inner method uses an explicit projection onto the vector complementarity set, which separates into independent scalar problems.

For the matrix case, Yan and Fukushima [39] proposed a smoothing method for symmetric-cone complementarity constraints, including the semidefinite case, based on a nonsmooth projection reformulation. Under suitable assumptions, limits of stationary points of the smoothed problems are C-stationary. For general SDCMPCCs, Wu and Zhang [35] studied a bilinear penalty framework and proved S-stationarity of feasible accumulation points under their assumptions. The resulting subproblems are nonlinear semidefinite programs, but their numerical solution is outside the scope of that analysis. For an inverse linear SDP, Wu et al. [37] developed a sequential convex programming method that exploits the DC structure of a nonsmooth partial penalty reformulation.

More recently, Santos et al. [32] analyze an augmented Lagrangian method based on a slack reformulation that retains semidefinite complementarity explicitly. They impose an approximate KKT condition on the separate semidefinite and scalar complementarity constraints. Their analysis gives AW-stationarity at feasible accumulation points and AC-stationarity under a multiplier-dependent accuracy requirement; a subsequent C-stationarity result uses an SDCMPCC-Robinson condition. These stationarity notions are defined in their formulation.

For second-order cone complementarity, Liang et al. [24] study an augmented Lagrangian method for a direct cone-and-scalar formulation. Its K-stationarity target is the KKT condition of that formulation. Ye and Zhou [40, Proposition 7.1 and Corollary 7.4] show that K-stationarity implies S-stationarity, while the converse can fail when a cone has dimension at least three. Thus a finite-multiplier target may impose an

additional representation requirement; our M-stationarity guarantee is not stronger than K-stationarity. Conic subproblems and positive-parameter relaxations are different objects from the unrelaxed scalar-orthogonality formulation.

For the semidefinite problem, the choice of a subproblem stationarity criterion depends on how complementarity is represented. Replacing Ω with two semidefinite constraints and one scalar orthogonality constraint preserves the feasible set, but the resulting finite multipliers need not represent all regular normals to Ω. This representation also has a nonzero abnormal multiplier at every feasible point, so convergence results requiring Robinson's constraint qualification cannot be invoked directly. Section 2.1 derives these properties, and Example 1 exhibits a missing coupled normal even at a strictly complementary point. These limitations motivate accuracy criteria based on the normal cone to Ω; they do not rule out conic algorithms.

For approaches that retain complementarity explicitly, the matrix projection is a central computational difficulty. The general augmented Lagrangian framework of Jia et al. [20] retains a closed set and penalizes the remaining smooth constraints. Its projected-gradient inner method requires projections onto the retained set. For semidefinite complementarity, a fixed common orthogonal basis reduces the projection problem to scalar problems with explicit solutions. Finding an optimal basis remains, in general, a coupled nonconvex optimization problem. Thus the inexpensive projection available for vector MPCCs does not directly extend to SDCMPCCs.

Building on this retained-set framework and the asymptotic stationarity analysis in [2, 29, 20], we develop SPALM for SDCMPCCs. Its slack candidates satisfy complementarity, and the outer method controls the smooth linking errors. Inexact proximal alternation produces candidates that must pass both descent and same-point limiting-normal tests. The stationarity target includes the coupled normals omitted by the scalar representation. Vanishing absolute errors yield AM-stationarity at feasible accumulation points, and AM-regularity permits passage to M-stationarity without assuming bounded raw multipliers for the degenerate scalar reformulation. Generating and certifying matrix candidates remains a computational cost. Section 3 compares the resulting accuracy requirement with the separate-constraint tests of [32]; no general cost or runtime ordering follows from the different interfaces.

The main contributions are as follows.

(i) We specialize the asymptotic stationarity framework to SPALM and establish AM-stationarity of feasible accumulation points of infinite outer sequences when the absolute subproblem errors vanish (Theorem 1). We characterize active-set SDCMPCC-LICQ and prove that it implies AM-regularity, which yields M-stationarity. Feasibility of accumulation points is treated separately from this stationarity implication.

(ii) To solve the augmented Lagrangian subproblem at each outer iteration, we develop an inexact proximal alternating framework with verifiable block descent and stationarity conditions. Boundedness of the initial sublevel set, two-sided bounds on the proximal parameters, and vanishing block errors yield convergence of the full residual to zero along infinite accepted sequences (Theorem 2). Square-summable block errors give an $O(T^{-1/2})$ best-residual bound over T accepted iterations (Corollary 3).

(iii) We construct Huber–Stiefel and fixed-basis procedures for generating feasible slack candidates, together with computable limiting-normal certificates for their verification. Acceptance requires both true descent and the prescribed normal-residual bound. We also derive spectral objective-gap bounds for the fixed-basis candidates.

(iv) We construct a compact rank lifting that preserves the projected feasible set and global optimal value of the rank-constrained correlation problem and satisfies active-set SDCMPCC-LICQ at every feasible point. The M-stationarity guarantees apply to this lifted formulation (Remark 7). For this lifting and a single-equality observation-fitting inverse SDP, we verify active-set SDCMPCC-LICQ and boundedness of the augmented Lagrangian sublevel sets under the stated model assumptions. For each fixed instance, comparing the inner initial point with an explicit feasible reference point gives a common upper bound on the outer AL values. With completed inner solves and vanishing absolute errors, this yields bounded primal iterates and M-stationary accumulation points (Lemma 2 and Corollary 5).

The numerical study evaluates SPALM on 24 parameter groups with five seeds per group. All 120 runs satisfy the prescribed numerical success tests, combining absolute feasibility and stationarity tests with a normalized witness-based residual target. Retained joint spectral data and independent numerical verification support the residual evaluations. The tables and figures summarize final residuals and solver wall-clock times. The finite-precision interpretation is given in Remark 6. The inverse SDP experiments assess fitting to an observed primal target included in the objective.

The remainder of the paper follows the analysis and computation outlined above. Section 2 develops the slack reformulation and stationarity framework. Section 3 presents SPALM and its outer convergence results. Section 4 analyzes accepted proximal alternating sequences, and Section 5 develops slack-candidate procedures and their verification. Section 6 presents the models and numerical results. The final section concludes the paper.

Notation. We denote by $\mathbb{S}^n$ the space of $n \times n$ real symmetric matrices, by $\mathbb{S}^n_+$ and $\mathbb{S}^n_-$ the cones of positive- and negative-semidefinite matrices, and by $\mathbb{O}^n$ the set of $n \times n$ orthogonal matrices. For $A, B \in \mathbb{R}^{n\times n}$, the Frobenius inner product is defined by $\langle A, B\rangle := \mathrm{tr}(A^\top B)$, and the corresponding Frobenius norm is $\|A\|_F := \sqrt{\langle A, A\rangle}$. The notation $\|\cdot\|$ denotes the norm induced by the given inner product on a Euclidean space; in particular, it is the Euclidean norm for vectors. For explicit matrix arguments, we write $\|\cdot\|_F$ for the Frobenius norm and $\|\cdot\|_2$ for the spectral norm. Product spaces are equipped with the canonical product norms induced by their component norms. In particular, for $u = (Y, Z) \in \mathbb{S}^n \times \mathbb{S}^n$, $\|u\|^2 = \|Y\|_F^2 + \|Z\|_F^2$. For a differentiable mapping F, $F'(w)$ and $F'(w)^*$ denote its derivative and adjoint derivative, respectively. For a scalar $s \in \mathbb{R}$, define $s_+ := \max\{s, 0\}$ and $s_- := \max\{-s, 0\}$. These operations are applied componentwise to vectors. Moreover, $\mathrm{Diag}(x)$ denotes the diagonal matrix generated by x, whereas $\mathrm{diag}(A)$ denotes the vector of diagonal entries of A. The eigenvalues $\lambda_i(A)$ follow the order in the stated spectral decomposition; $\lambda_j^{\downarrow}(A)$ denotes the j-th largest eigenvalue. Any spectral reordering also reorders the corresponding eigenvector columns. For a nonempty closed convex set D, Π_D denotes its Euclidean projection, and

$$N_D(x) := \{v : \langle v, z - x\rangle \le 0 \text{ for every } z \in D\} \quad (x \in D), \qquad N_D(x) := \varnothing \quad (x \notin D).$$

Regular and limiting normals to closed nonconvex sets are defined below.

2 Problem reformulation and stationarity framework

We separate the smooth constraints from semidefinite complementarity, establish M-stationarity equivalence under the unique slack lift, and introduce the asymptotic stationarity framework used by SPALM. Active-set SDCMPCC-LICQ and the absence of singular M-multipliers provide sufficient conditions for the required AM-regularity.

2.1 Slack reformulation and normal-cone geometry

To fit problem (1) into the structured geometric constraint framework of [20], we introduce $u = (Y, Z) \in \mathcal{U} := \mathbb{S}^n \times \mathbb{S}^n$ and $x = (w, u) \in \mathcal{X} := \mathcal{Z} \times \mathcal{U}$. Problem (1) can be equivalently reformulated as

$$\min_{w,Y,Z} f(w) \quad \text{s.t.} \quad g(w) \le 0, \quad h(w) = 0, \quad G(w) = Y, \quad H(w) = Z, \quad (Y, Z) \in \Omega. \tag{2}$$

Set $\varphi(x) := f(w)$ and

$$\Psi(x) := (g(w), h(w), G(w) - Y, H(w) - Z), \qquad C := \mathbb{R}^p_- \times \{0\}^q \times \{0\}^{\mathbb{S}^n} \times \{0\}^{\mathbb{S}^n}.$$

Thus (2) minimizes $\varphi(x)$ subject to $\Psi(x) \in C$ and $x \in \mathcal{Z} \times \Omega$. Write

$$\mathcal{F} := \{w : g(w) \le 0,\ h(w) = 0,\ (G(w), H(w)) \in \Omega\}, \qquad \widetilde{\mathcal{F}} := \{x \in \mathcal{Z} \times \Omega : \Psi(x) \in C\}.$$

The map $w \mapsto (w, (G(w), H(w)))$ is a bijection from $\mathcal{F}$ to $\widetilde{\mathcal{F}}$ and preserves objective values. In particular, the optimal values of the two formulations coincide. For any $u = (Y, Z) \in \Omega$, the linking error also bounds the distance of the matrix image to complementarity:

$$\mathrm{dist}((G(w), H(w)), \Omega) \le \left(\|G(w) - Y\|_F^2 + \|H(w) - Z\|_F^2\right)^{1/2}. \tag{3}$$

We recall the following standard notions from variational analysis [31]. For a closed subset S of a finite-dimensional Euclidean space, $x^k \to_S \bar{x}$ means $x^k \in S$ and $x^k \to \bar{x}$. The outer limit is defined by $\mathrm{Limsup}_{x\to\bar{x}} \mathcal{T}(x) := \{v : \exists\, x^k \to \bar{x},\ v^k \to v,\ v^k \in \mathcal{T}(x^k) \text{ for all } k\}$.

Definition 1 (Regular and limiting normal cones) For $\bar{x} \in S$, define

$$\widehat{N}_S(\bar{x}) := \left\{ v : \limsup_{\substack{x \to_S \bar{x} \\ x \neq \bar{x}}} \frac{\langle v, x - \bar{x}\rangle}{\|x - \bar{x}\|} \le 0 \right\}, \qquad N_S^{\lim}(\bar{x}) := \operatorname*{Limsup}_{x \to_S \bar{x}} \widehat{N}_S(x).$$

Both cones are empty outside S. We have $\widehat{N}_S \subseteq N_S^{\lim}$, and the limiting normal-cone mapping $N_S^{\lim}$ is outer semicontinuous.

For $u = (Y, Z) \in \Omega$, consider the common spectral representation [9, Sections 2.3 and 3]:

$$\mathcal{E}(u) := Y + Z = P\,\mathrm{Diag}(\Lambda_{\alpha\alpha}, 0, \Lambda_{\gamma\gamma})P^\top, \qquad \Lambda_{\alpha\alpha} \succ 0, \quad \Lambda_{\gamma\gamma} \prec 0. \tag{4}$$

Here α, β, γ are the positive, zero, and negative spectral index sets; $P = (P_\alpha, P_\beta, P_\gamma) \in \mathbb{O}^n$, $Y = P_\alpha \Lambda_{\alpha\alpha} P_\alpha^\top$, and $Z = P_\gamma \Lambda_{\gamma\gamma} P_\gamma^\top$.

$$\Sigma_{ij} := \frac{\lambda_i(\mathcal{E}(u))}{\lambda_i(\mathcal{E}(u)) - \lambda_j(\mathcal{E}(u))} \quad (i \in \alpha,\ j \in \gamma), \qquad \widetilde{\Gamma}^Y := P^\top \Gamma^Y P, \quad \widetilde{\Gamma}^Z := P^\top \Gamma^Z P. \tag{5}$$

Let E_{IJ} denote the all-ones matrix on $I \times J$, and let $\circ$ denote the Hadamard product. The common multiplier space is

$$\mathcal{K}_W(u) := \Big\{(\Gamma^Y, \Gamma^Z) \in \mathbb{S}^n \times \mathbb{S}^n : \widetilde{\Gamma}^Y_{\alpha\alpha} = \widetilde{\Gamma}^Y_{\alpha\beta} = \widetilde{\Gamma}^Y_{\beta\alpha} = 0, \quad \widetilde{\Gamma}^Z_{\gamma\gamma} = \widetilde{\Gamma}^Z_{\beta\gamma} = \widetilde{\Gamma}^Z_{\gamma\beta} = 0,$$
$$\Sigma_{\alpha\gamma} \circ \widetilde{\Gamma}^Y_{\alpha\gamma} + (E_{\alpha\gamma} - \Sigma_{\alpha\gamma}) \circ \widetilde{\Gamma}^Z_{\alpha\gamma} = 0\Big\}.$$

Multiplication in the basis P gives the basis-independent identity

$$(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) \quad \Longleftrightarrow \quad Y\Gamma^Y - \Gamma^Z Z = 0. \tag{6}$$

Indeed, the $\alpha\alpha$ and $\alpha\beta$ blocks of $\widetilde{\Gamma}^Y$ and the $\beta\gamma$ and $\gamma\gamma$ blocks of $\widetilde{\Gamma}^Z$ vanish by invertibility of the nonzero spectral blocks. For $i \in \alpha$ and $j \in \gamma$, the remaining relation is $\lambda_i(\mathcal{E}(u))\widetilde{\Gamma}^Y_{ij} - \lambda_j(\mathcal{E}(u))\widetilde{\Gamma}^Z_{ij} = 0$, which gives the $\alpha\gamma$ condition by (5). A simultaneous rotation of the positive and negative eigenspaces gives a geometric interpretation of this coupling in dimension two.

Example 1 (Coupled normals at a strictly complementary pair) Let $n = 2$ and set

$$Y_0 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad Z_0 = -\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad W = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

The pair $u_0 = (Y_0, Z_0)$ is strictly complementary since $Y_0 - Z_0 = I_2$. With $p_\theta = (\cos\theta, \sin\theta)^\top$ and $q_\theta = (-\sin\theta, \cos\theta)^\top$, every feasible pair near u_0 has the form $(ap_\theta p_\theta^\top, -bq_\theta q_\theta^\top)$ for $a, b > 0$. Differentiating this local parameterization at $(a, b, \theta) = (1, 1, 0)$ gives

$$T_\Omega(u_0) = \mathrm{span}\{(Y_0, 0), (0, Z_0), (W, W)\}.$$

Orthogonality to the rotation direction (W, W) imposes $\Gamma^Y_{12} + \Gamma^Z_{12} = 0$ on each normal pair. Taking the orthogonal complement of the tangent space of this local manifold gives

$$\widehat{N}_\Omega(u_0) = N^{\lim}_\Omega(u_0) = \left\{ \left(\begin{pmatrix} 0 & r \\ r & s \end{pmatrix}, \begin{pmatrix} t & -r \\ -r & 0 \end{pmatrix} \right) : r, s, t \in \mathbb{R} \right\}. \tag{7}$$

To compute the finite multiplier image at the same pair, write the scalar representation in its general form $\Omega = F^{-1}(D)$, where $F(Y, Z) = (Y, Z, \langle Y, Z\rangle)$ and $D = \mathbb{S}^n_+ \times \mathbb{S}^n_- \times \mathbb{R}_+$. On $\mathbb{S}^n_+ \times \mathbb{S}^n_-$, the inner product $\langle Y, Z\rangle$ is nonpositive and vanishes exactly when $YZ = 0$.[1] At $u = (Y, Z) \in \Omega$, the finite multiplier image is

$$\begin{aligned} \mathcal{N}^{\mathrm{img}}(Y, Z) &:= F'(u)^* N_D(F(u)) \\ &= \big(N_{\mathbb{S}^n_+}(Y) \times N_{\mathbb{S}^n_-}(Z)\big) + \mathbb{R}_-(Z, Y) \subseteq \widehat{N}_\Omega(u). \end{aligned} \tag{9}$$

For $u + (\Delta Y, \Delta Z) \in \Omega$, orthogonality at both feasible pairs gives $\langle Z, \Delta Y\rangle + \langle Y, \Delta Z\rangle = -\langle \Delta Y, \Delta Z\rangle = O(\|(\Delta Y, \Delta Z)\|^2)$. Together with the two cone-normal inequalities, this proves the inclusion in (9).

At u_0, the two cone normals are $N_{\mathbb{S}^n_+}(Y_0) = \mathbb{R}_+ Z_0$ and $N_{\mathbb{S}^n_-}(Z_0) = \mathbb{R}_+ Y_0$. Hence the finite multiplier image consists of the diagonal pairs

$$\mathcal{N}^{\mathrm{img}}(Y_0, Z_0) = \{(sZ_0, tY_0) : s, t \in \mathbb{R}\}.$$

The scalar equality $\langle Y, Z\rangle = 0$ gives the same image at u_0. The off-diagonal pair $\xi_0 = (W, -W)$ belongs to the normal space in (7) and is orthogonal to this multiplier image. Since $\|\xi_0\| = 2$, the two distances are

$$\mathrm{dist}\big(\xi_0, N^{\lim}_\Omega(u_0)\big) = 0, \qquad \mathrm{dist}\big(\xi_0, \mathcal{N}^{\mathrm{img}}(Y_0, Z_0)\big) = 2. \tag{10}$$

[1] At every feasible $u = (Y, Z)$, the tuple $(Z, Y, -1)$ satisfies

$$(Z, Y, -1) \in N_D(F(u)) \setminus \{0\}, \qquad F'(u)^*(Z, Y, -1) = 0. \tag{8}$$

After composition with (G, H), this direction certifies the failure of Robinson's constraint qualification for the scalar representation of (1) [9, Proposition 4.1]. Adding its nonnegative multiples to any KKT multiplier gives an unbounded ray of multipliers.

A path with derivative ξ_0 illustrates the local scaling of the scalar orthogonality residual. Consider the positive and negative semidefinite matrices

$$Y(t) = \begin{pmatrix} 1 & t \\ t & t^2 \end{pmatrix}, \qquad Z(t) = -\begin{pmatrix} t^2 & t \\ t & 1 \end{pmatrix}.$$

Along this path, $|\langle Y(t), Z(t)\rangle| = 4t^2$. Since $(Y'(0), Z'(0)) = \xi_0$ is normal to the local manifold at u_0 and has norm 2, the distance to Ω satisfies $\operatorname{dist}((Y(t), Z(t)), \Omega) = 2|t| + o(|t|)$ as $t \to 0$.

In Example 1, the common kernel is trivial and the $\alpha\gamma$ coupling is already present. When Y and Z have a nontrivial common kernel, the conditions on the $\beta\beta$ blocks distinguish the following spectral multiplier sets [9].

Definition 2 (Spectral multiplier sets) For $u \in \Omega$, define

$$\begin{aligned} \mathcal{K}_S(u) &:= \{(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : \widetilde{\Gamma}^Y_{\beta\beta} \preceq 0,\ \widetilde{\Gamma}^Z_{\beta\beta} \succeq 0\}, \\ \mathcal{K}_M(u) &:= N^{\lim}_{\Omega}(u), \\ \mathcal{K}_C(u) &:= \{(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : \langle \widetilde{\Gamma}^Y_{\beta\beta}, \widetilde{\Gamma}^Z_{\beta\beta}\rangle \le 0\}. \end{aligned}$$

Appendix A recalls the full spectral characterization of $\mathcal{K}_M(u)$.

These sets satisfy

$$\mathcal{K}_S(u) = \widehat{N}_{\Omega}(u) \subseteq \mathcal{K}_M(u) \subseteq \mathcal{K}_C(u) \subseteq \mathcal{K}_W(u), \qquad \operatorname{span} \widehat{N}_{\Omega}(u) = \operatorname{span} N^{\lim}_{\Omega}(u) = \mathcal{K}_W(u).$$

The span identities follow from the inclusion chain and the fact that every symmetric $\beta\beta$ block is a difference of two positive semidefinite matrices. The remaining block relations are homogeneous and independent of the $\beta\beta$ blocks.

2.2 Stationarity and asymptotic stationarity

Using the spectral multiplier sets introduced above and the stationarity framework of [9], we define the S-, M-, C-, and W-stationarity conditions and describe their relationships.

Definition 3 (M-stationarity for the SDCMPCC) A feasible point $\bar{w} \in \mathcal{F}$, with $\bar{u} = (G(\bar{w}), H(\bar{w}))$, is M-stationary if there exist $\lambda \in \mathbb{R}^p$, $\mu \in \mathbb{R}^q$, and $\Gamma^Y, \Gamma^Z \in \mathbb{S}^n$ such that

$$0 = \nabla f(\bar{w}) + g'(\bar{w})^*\lambda + h'(\bar{w})^*\mu + G'(\bar{w})^*\Gamma^Y + H'(\bar{w})^*\Gamma^Z, \tag{11}$$

$$\lambda \ge 0, \qquad \langle \lambda, g(\bar{w})\rangle = 0, \tag{12}$$

and

$$(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_M(\bar{u}). \tag{13}$$

For (2), M-stationarity at $(\bar{w}, \bar{u}) \in \widetilde{\mathcal{F}}$ involves additional linking multipliers $M^G, M^H \in \mathbb{S}^n$ satisfying

$$0 = \nabla f(\bar{w}) + g'(\bar{w})^*\lambda + h'(\bar{w})^*\mu + G'(\bar{w})^*M^G + H'(\bar{w})^*M^H, \tag{14}$$

$$M^G = \Gamma^Y, \qquad M^H = \Gamma^Z, \tag{15}$$

together with (12)–(13).

The following equivalence allows stationarity results for the slack reformulation to be transferred to the original problem.

Proposition 1 (Equivalence of M-stationarity) *For $\bar{w} \in \mathcal{F}$ and $\bar{u} = (G(\bar{w}), H(\bar{w}))$, $\bar{w}$ is M-stationary for (1) if and only if $(\bar{w}, \bar{u})$ is M-stationary for (2).*

Proof Eliminating M^G, M^H from (14)–(15) yields (11). Conversely, setting $M^G = \Gamma^Y$ and $M^H = \Gamma^Z$ gives the M-stationarity conditions for (2). □

Replacing $\mathcal{K}_M$ in (13) by $\mathcal{K}_S$, $\mathcal{K}_C$, or $\mathcal{K}_W$ defines S-, C-, or W-stationarity, respectively. The inclusions among these multiplier sets yield

$$\text{S-stationarity} \implies \text{M-stationarity} \implies \text{C-stationarity} \implies \text{W-stationarity}$$

We next specialize the asymptotic stationarity and regularity framework of [20, Section 2.2] to the slack reformulation (2). Define the multiplier image

$$\mathcal{M}(w,u,z) := \Psi'(w,u)^* N_C(\Psi(w,u)-z) + (\{0\} \times N_\Omega^{\lim}(u)), \tag{16}$$

where $w \in \mathcal{Z}$, $u \in \mathcal{U}$, and $z \in \mathbb{R}^p \times \mathbb{R}^q \times \mathbb{S}^n \times \mathbb{S}^n$. Since $\nabla\varphi(w,u) = (\nabla f(w), 0, 0)$ and $N_{\mathcal{Z}\times\Omega}^{\lim}(w,u) = \{0\} \times N_\Omega^{\lim}(u)$, exact M-stationarity is

$$0 \in \nabla\varphi(\bar{w},\bar{u}) + \mathcal{M}(\bar{w},\bar{u},0).$$

Definition 4 (AM-stationarity and AM-regularity) Let $(\bar{w},\bar{u}) \in \widetilde{\mathcal{F}}$.

(i) The point is asymptotically M-stationary (AM-stationary) if there exist $(w^k,u^k) \to (\bar{w},\bar{u})$, $z^k \to 0$, and $e^k \to 0$ with $e^k \in \nabla\varphi(w^k,u^k) + \mathcal{M}(w^k,u^k,z^k)$ for every k.

(ii) Problem (2) is AM-regular at this point if

$$\operatorname*{Limsup}_{(w,u,z)\to(\bar{w},\bar{u},0)} \mathcal{M}(w,u,z) \subseteq \mathcal{M}(\bar{w},\bar{u},0).$$

Condition (i) requires $u^k \in \Omega$ and $\Psi(w^k,u^k) - z^k \in C$, while (w^k,u^k) need not be feasible for (2). AM-regularity ensures the outer semicontinuity of the multiplier image, under which AM-stationarity implies M-stationarity.

2.3 Active-set SDCMPCC-LICQ

To establish AM-regularity for the slack reformulation, we adapt the spectral SDCMPCC-LICQ condition in [9, Theorem 6.1(ii)] to the active inequality constraints and derive an equivalent local uniform estimate.

For $\Upsilon = (\lambda,\mu,M^G,M^H)$ and $\Gamma = (\Gamma^Y,\Gamma^Z)$, define

$$\mathcal{J}_w(\Upsilon,\Gamma) := \Psi'(w,u)^*\Upsilon + (0,\Gamma) = \begin{pmatrix} g'(w)^*\lambda + h'(w)^*\mu + G'(w)^*M^G + H'(w)^*M^H \\ -M^G + \Gamma^Y \\ -M^H + \Gamma^Z \end{pmatrix}. \tag{17}$$

Call (w,u,z) admissible if $u \in \Omega$ and $s := \Psi(w,u) - z \in C$. For such a triple, set $\mathcal{I}(s_g) := \{i : (s_g)_i = 0\}$ and

$$\mathfrak{L}(w,u,z) := \operatorname{span} N_C(s) \times \mathcal{K}_W(u).$$

Lemma 1 (Closedness of the linearized multiplier space) *For an admissible triple with $u = (Y,Z)$,*

$$\mathfrak{L}(w,u,z) = \{(\lambda,\mu,M^G,M^H,\Gamma^Y,\Gamma^Z) : \operatorname{Diag}(s_g)\lambda = 0, \quad Y\Gamma^Y - \Gamma^Z Z = 0\}. \tag{18}$$

The mapping $\mathfrak{L}$ has a closed graph on admissible triples.

Proof The activity condition is exactly $\operatorname{Diag}(s_g)\lambda = 0$, and (6) gives the spectral condition. These equalities are preserved under limits, as is admissibility by continuity of Ψ and closedness of C and Ω. □

Definition 5 (Active-set SDCMPCC-LICQ) Active-set SDCMPCC-LICQ holds at $(\bar{w},\bar{u}) \in \widetilde{\mathcal{F}}$ if $\ker \mathcal{J}_{\bar{w}} \cap \mathfrak{L}(\bar{w},\bar{u},0) = \{0\}$.

Proposition 2 (Reduced and uniform characterizations of LICQ) *For $(\bar{w},\bar{u}) \in \widetilde{\mathcal{F}}$, with $\bar{u} = (\bar{Y},\bar{Z})$, the following are equivalent:*

(i) Active-set SDCMPCC-LICQ holds at $(\bar{w},\bar{u})$.

(ii) The only tuple $(\lambda,\mu,\Gamma^Y,\Gamma^Z)$ satisfying

$$g'(\bar{w})^*\lambda + h'(\bar{w})^*\mu + G'(\bar{w})^*\Gamma^Y + H'(\bar{w})^*\Gamma^Z = 0,$$
$$\lambda_i = 0 \quad (i \notin \mathcal{I}(g(\bar{w}))), \qquad \bar{Y}\Gamma^Y - \Gamma^Z\bar{Z} = 0$$

is zero.

(iii) There exist a neighborhood $\mathcal{N}$ of $(\bar{w},\bar{u},0)$ and $c > 0$ such that

$$\|\mathcal{J}_w t\| \geq c\|t\| \quad \text{for every admissible } (w,u,z) \in \mathcal{N}, \quad t \in \mathfrak{L}(w,u,z). \tag{19}$$

Proof The last two equations of $\mathcal{J}_{\bar{w}}t = 0$ give $M^G = \Gamma^Y$ and $M^H = \Gamma^Z$; elimination proves (i)⇔(ii). If (iii) fails, choose admissible triples within $1/k$ of $(\bar{w}, \bar{u}, 0)$ and nonzero tuples with $\|\mathcal{J}_{w^k} t^k\| < \|t^k\|/k$. Normalize them and pass to a convergent subsequence on the unit sphere. Lemma 1 and continuity of $\mathcal{J}_w$ give a unit tuple in $\ker \mathcal{J}_{\bar{w}} \cap \mathfrak{L}(\bar{w}, \bar{u}, 0)$, contradicting (i). Conversely, (iii) applied at the base point implies (i). □

Proposition 3 (AM-regularity under active-set SDCMPCC-LICQ) *If active-set SDCMPCC-LICQ holds at $(\bar{w}, \bar{u}) \in \widetilde{\mathcal{F}}$, then (2) is AM-regular there.*

Proof Let admissible triples $(w^k, u^k, z^k) \to (\bar{w}, \bar{u}, 0)$ and $v^k \in \mathcal{M}(w^k, u^k, z^k)$ satisfy $v^k \to \bar{v}$. Choose representations $v^k = \mathcal{J}_{w^k} t^k$, where

$$t^k = (\Upsilon^k, \Gamma^k) \in N_C(\Psi(w^k, u^k) - z^k) \times N_\Omega^{\lim}(u^k) \subseteq \mathfrak{L}(w^k, u^k, z^k).$$

The uniform estimate (19) gives $c\|t^k\| \le \|v^k\|$ for all sufficiently large k. Thus, after passing to a subsequence, $t^k \to (\bar{\Upsilon}, \bar{\Gamma})$. By the closed-graph properties of the normal-cone mappings, $(\bar{\Upsilon}, \bar{\Gamma}) \in N_C(\Psi(\bar{w}, \bar{u})) \times N_\Omega^{\lim}(\bar{u})$. Continuity then gives $\bar{v} = \mathcal{J}_{\bar{w}}(\bar{\Upsilon}, \bar{\Gamma}) \in \mathcal{M}(\bar{w}, \bar{u}, 0)$. □

Remark 1 (NNAMCQ and AM-regularity) A weaker sufficient condition for AM-regularity is the absence of singular M-multipliers. Following [9, Definition 6.1], a nonzero tuple $(\lambda, \mu, \Gamma^Y, \Gamma^Z)$ is called a *singular M-multiplier* for (1) at $\bar{w} \in \mathcal{F}$ if

$$g'(\bar{w})^*\lambda + h'(\bar{w})^*\mu + G'(\bar{w})^*\Gamma^Y + H'(\bar{w})^*\Gamma^Z = 0,$$
$$\lambda \in N_{\mathbb{R}^p_-}(g(\bar{w})), \qquad (\Gamma^Y, \Gamma^Z) \in N_\Omega^{\lim}(\bar{u}),$$

where $\bar{u} = (G(\bar{w}), H(\bar{w}))$. The absence of such multipliers is equivalent to the no nonzero abnormal multiplier constraint qualification (NNAMCQ) for (2), since the linking multipliers satisfy $M^G = \Gamma^Y$ and $M^H = \Gamma^Z$.

Since

$$N_C(\Psi(\bar{w}, \bar{u})) \times N_\Omega^{\lim}(\bar{u}) \subseteq \mathfrak{L}(\bar{w}, \bar{u}, 0),$$

active-set SDCMPCC-LICQ implies NNAMCQ, which in turn implies AM-regularity by [20, Section 2.2, Eq. (2.2)].

The first implication is strict. Let $n = 1$, $G(w) = w$, and $H(w) = -w$, with no other constraints. At the unique feasible point $w = 0$, the homogeneous stationarity equation is $\Gamma^Y = \Gamma^Z$. While $\mathcal{K}_W(0, 0) = \mathbb{R}^2$ admits the nonzero solution $(1, 1)$, we have

$$\{(\Gamma^Y, \Gamma^Z) \in N_\Omega^{\lim}(0, 0) : \Gamma^Y = \Gamma^Z\} = \{(0, 0)\}.$$

Hence, active-set SDCMPCC-LICQ fails, whereas NNAMCQ holds.

Remark 2 (Scalar complementarity specialization) For $\Omega_{\mathrm{cc}} = \operatorname{gph} N_{\mathbb{R}^m_+}$, write $y = G(w)$ and $\zeta = -H(w)$. The usual nonnegative pair (y, ζ) defines I^{+0}, I^{0+}, I^{00}, and

$$\mathcal{K}_W^{\mathrm{cc}}(y, -\zeta) = \{(\Gamma^Y, \Gamma^Z) : \Gamma_i^Y = 0 \ (i \in I^{+0}),\ \Gamma_i^Z = 0 \ (i \in I^{0+})\}.$$

In this scalar setting, the active-set SDCMPCC-LICQ reduces to the classical MPCC-LICQ [4, Definition 2], with Γ^Z corresponding to $H = -\zeta$. However, diagonal-valued maps into the full matrix cone need not yield the same LICQ, since off-diagonal multipliers remain admissible.

3 Structure-preserving augmented Lagrangian method

SPALM penalizes the smooth constraints of (2) and retains Ω explicitly. We separate feasibility of accumulation points from their AM-stationarity; AM-regularity then gives M-stationarity.

3.1 Augmented Lagrangian function

Following the partial-penalization strategy in [15,20], we incorporate $\Psi(x) \in C$ into the objective and retain $x \in \mathcal{Z} \times \Omega$.

For $\rho > 0$ and $\Upsilon = (\lambda, \mu, M^G, M^H)$, define the augmented Lagrangian function by

$$\mathcal{L}_\rho(x; \Upsilon) := \varphi(x) + \frac{\rho}{2} \operatorname{dist}^2 \left(\Psi(x) + \frac{\Upsilon}{\rho}, C \right).$$

Using the product structure of C, we obtain the following expression in $x = (w, Y, Z)$:

$$\begin{aligned} \mathcal{L}_\rho(w, Y, Z; \lambda, \mu, M^G, M^H) = f(w) &+ \frac{\rho}{2} \left\| \left(g(w) + \frac{\lambda}{\rho} \right)_+ \right\|^2 + \frac{\rho}{2} \left\| h(w) + \frac{\mu}{\rho} \right\|^2 \\ &+ \frac{\rho}{2} \left\| G(w) - Y + \frac{M^G}{\rho} \right\|_F^2 + \frac{\rho}{2} \left\| H(w) - Z + \frac{M^H}{\rho} \right\|_F^2. \end{aligned} \tag{20}$$

At the k-th outer iteration, let $\bar{\Upsilon}^k := (\bar{\lambda}^k, \bar{\mu}^k, \bar{M}^{G,k}, \bar{M}^{H,k})$ be the safeguarded multiplier vector, and define

$$L_k(x) := \mathcal{L}_{\rho_k}(x; \bar{\Upsilon}^k).$$

The corresponding augmented Lagrangian subproblem is

$$\min_{x \in \mathcal{Z} \times \Omega} \; L_k(x). \tag{21}$$

Step 3 below requires approximate M-stationarity for this subproblem.

3.2 SPALM for SDCMPCCs

The penalty update and stopping test use a common feasibility measure. For $x = (w, u)$, with $u = (Y, Z)$, $\lambda \in \mathbb{R}^p$, and $\rho > 0$, define

$$\mathcal{V}_\rho(x; \lambda) := \max \left\{ \left\| g(w) - \Pi_{\mathbb{R}^p_-} \left(g(w) + \frac{\lambda}{\rho} \right) \right\|, \|h(w)\|, \|G(w) - Y\|_F, \|H(w) - Z\|_F \right\}. \tag{22}$$

This projection-based measure bounds the perturbation in the smooth constraints and in the normal-cone relation for the updated multipliers, as shown below. The first term vanishes when the inequality constraints and the corresponding multiplier conditions are satisfied.

Feasible-reference comparison. When a fixed feasible point $x^{\mathrm{ref}} = (w^{\mathrm{ref}}, u^{\mathrm{ref}})$ of (2) is available, we use the following optional rule in Step 3:

$$x^{k,0} \in \operatorname*{arg\,min}_{x \in \{x^k, x^{\mathrm{ref}}\}} L_k(x), \qquad L_k(x^{k+1}) \le L_k(x^{k,0}). \tag{23}$$

The reference point remains fixed across outer iterations, and both AL values are evaluated with the current parameters $(\rho_k, \bar{\Upsilon}^k)$. Accepted inner descent ensures the second inequality in (23).

Algorithm 1 Structure-preserving augmented Lagrangian method (SPALM)

Input: An initial point $x^0 = (w^0, u^0) \in \mathcal{Z} \times \Omega$, with $u^0 = (Y^0, Z^0)$; initial multipliers $\Upsilon^0 = (\lambda^0, \mu^0, M^{G,0}, M^{H,0})$; a penalty parameter $\rho_0 > 0$; parameters $\beta_\rho > 1$ and $\eta \in (0,1)$; safeguards $\lambda_{\max} > 0$, $\mu_{\min} < \mu_{\max}$, $M^G_{\max} > 0$, and $M^H_{\max} > 0$; a feasibility tolerance $\varepsilon_{\text{feas}} > 0$; a stationarity tolerance $\varepsilon_{\text{stat}} > 0$; and a nonincreasing positive sequence $\{\nu_k\}$ satisfying $\nu_k \to 0$.

Output: A computed iterate $x^k = (w^k, u^k)$, with $u^k = (Y^k, Z^k)$, its updated multipliers Υ^k, and a normal certificate for the slack stationarity test.

Set $k = 0$ and $V_0 := \mathcal{V}_{\rho_0}(x^0; 0)$.

Step 1: feasibility and stationarity termination test.

If $k \geq 1$ and

$$V_k \leq \varepsilon_{\text{feas}}, \qquad \left\|\nabla_w L_{k-1}(x^k)\right\| \leq \varepsilon_{\text{stat}}, \qquad \operatorname{dist}\left(-\nabla_u L_{k-1}(x^k), N^{\lim}_{\Omega}(u^k)\right) \leq \varepsilon_{\text{stat}},$$

output x^k and Υ^k, retain a normal certificate for the slack test, and terminate.

Step 2: safeguard the multipliers.

Project the vector multipliers onto their prescribed boxes and the matrix multipliers onto Frobenius balls:

$$\bar{\lambda}^k := \Pi_{[0,\lambda_{\max}]^p}(\lambda^k), \qquad \bar{\mu}^k := \Pi_{[\mu_{\min},\mu_{\max}]^q}(\mu^k).$$

$$\bar{M}^{G,k} := \frac{M^{G,k}}{\max\{1, \|M^{G,k}\|_F / M^G_{\max}\}}, \qquad \bar{M}^{H,k} := \frac{M^{H,k}}{\max\{1, \|M^{H,k}\|_F / M^H_{\max}\}}.$$

Set $\bar{\Upsilon}^k := (\bar{\lambda}^k, \bar{\mu}^k, \bar{M}^{G,k}, \bar{M}^{H,k})$.

Step 3: approximately solve the AL subproblem.

If x^{ref} is supplied, use (23).

Compute $x^{k+1} = (w^{k+1}, u^{k+1}) \in \mathcal{Z} \times \Omega$, with $u^{k+1} = (Y^{k+1}, Z^{k+1})$ such that

$$\left\|\nabla_w L_k(x^{k+1})\right\| \leq \nu_{k+1}, \qquad \operatorname{dist}\left(-\nabla_u L_k(x^{k+1}), N^{\lim}_{\Omega}(u^{k+1})\right) \leq \nu_{k+1}.$$

Step 4: update the multipliers.

Set

$$\lambda^{k+1} := \left(\bar{\lambda}^k + \rho_k g(w^{k+1})\right)_+, \qquad \mu^{k+1} := \bar{\mu}^k + \rho_k h(w^{k+1}),$$

$$M^{G,k+1} := \bar{M}^{G,k} + \rho_k\big(G(w^{k+1}) - Y^{k+1}\big), \qquad M^{H,k+1} := \bar{M}^{H,k} + \rho_k\big(H(w^{k+1}) - Z^{k+1}\big).$$

Step 5: compute the new feasibility measure.

Set

$$V_{k+1} := \mathcal{V}_{\rho_k}(x^{k+1}; \bar{\lambda}^k).$$

Step 6: update the penalty parameter.

Set

$$\rho_{k+1} := \begin{cases} \rho_k, & \text{if } V_{k+1} \leq \eta V_k, \\ \beta_\rho \rho_k, & \text{otherwise.} \end{cases}$$

Set $k \leftarrow k + 1$ and return to Step 1.

Multiplier relations. The safeguards give $M_{\text{saf}} := \sup_k \|\bar{\Upsilon}^k\| < \infty$, while the updated multipliers Υ^{k+1} may be unbounded. To express the multiplier update and the subproblem gradient, define

$$z^{k+1} := \Psi(x^{k+1}) - \Pi_C\left(\Psi(x^{k+1}) + \frac{\bar{\Upsilon}^k}{\rho_k}\right). \tag{24}$$

Projection optimality and the multiplier update give

$$\Upsilon^{k+1} = \bar{\Upsilon}^k + \rho_k z^{k+1} \in N_C\big(\Psi(x^{k+1}) - z^{k+1}\big). \tag{25}$$

Differentiating the squared-distance expression for L_k gives

$$\nabla L_k(x^{k+1}) = \nabla\varphi(x^{k+1}) + \Psi'(x^{k+1})^* \Upsilon^{k+1}. \tag{26}$$

In particular, $\nabla_u L_k(x^{k+1}) = -(M^{G,k+1}, M^{H,k+1})$, so the slack test in Step 3 measures the distance of the updated linking multipliers to $N^{\lim}_{\Omega}(u^{k+1})$.

Perturbation bounds. The four blocks of z^{k+1} are the residuals in (22); hence

$$V_{k+1} \leq \|z^{k+1}\| \leq 2V_{k+1}, \qquad \operatorname{dist}\big(\Psi(x^{k+1}), C\big) \leq \|z^{k+1}\|. \tag{27}$$

Since projection onto C does not increase distances, we also obtain

$$\|z^{k+1}\| \leq \operatorname{dist}\big(\Psi(x^{k+1}), C\big) + \frac{\|\bar{\Upsilon}^k\|}{\rho_k} \leq \operatorname{dist}\big(\Psi(x^{k+1}), C\big) + \frac{M_{\text{saf}}}{\rho_k}. \tag{28}$$

Termination and error bounds. Suppose that Step 1 terminates at x^k. Choose $\Gamma^k \in N^{\lim}_{\Omega}(u^k)$ satisfying $\|\nabla_u L_{k-1}(x^k) + \Gamma^k\| \leq \varepsilon_{\text{stat}}$; such a normal exists because $N^{\lim}_{\Omega}(u^k)$ is nonempty and closed. A retained inner

normal with residual at most ν_k can be reused when $\nu_k \le \varepsilon_{\mathrm{stat}}$; cf. [20, equations (4.4)–(4.5)]. With z^k defined by (24) and $e^k := \nabla L_{k-1}(x^k) + (0, \Gamma^k)$, the projection and gradient identities, together with the stopping test, give

$$\begin{aligned} x^k \in \mathcal{Z} \times \Omega, \qquad & \Psi(x^k) - z^k \in C, \\ \Upsilon^k \in N_C\big(\Psi(x^k) - z^k\big), \qquad & \Gamma^k \in N_\Omega^{\lim}(u^k), \\ e^k = \nabla\varphi(x^k) + \Psi'(x^k)^*\Upsilon^k & + (0, \Gamma^k), \\ \|z^k\| \le 2\varepsilon_{\mathrm{feas}}, \qquad & \|e^k\| \le \sqrt{2}\,\varepsilon_{\mathrm{stat}}. \end{aligned} \tag{29}$$

Thus the returned point satisfies the perturbed M-stationarity system (29) with the stated residual bounds. For the inequality block z_g^k, the normal relation yields $\lambda_i^k g_i(w^k) = \lambda_i^k z_{g,i}^k$ for $i = 1, \dots, p$, so bounds on the unperturbed complementarity products also depend on the magnitudes of the updated multipliers.

3.3 Global convergence analysis

We first establish sufficient conditions for feasibility of accumulation points and then analyze their AM- and M-stationarity. For the slack reformulation (2), the objective is $\varphi(x) = f(w)$, the smooth constraint mapping is Ψ, and the retained set is $\mathcal{Z} \times \Omega$.

Jia et al. [20, Proposition 4.1] gave sufficient conditions for feasibility of accumulation points. Our feasibility measure (22) uses the maximum of the four residual block norms. In view of the norm equivalence in (27), their result yields the following proposition.

Proposition 4 (Feasibility of accumulation points) *Suppose that Algorithm 1 generates an infinite sequence $\{x^{k+1}\}$. If either of the following conditions holds, then every accumulation point of this sequence is feasible for* (2)*:*

(i) The sequence $\{\rho_k\}$ is bounded.
(ii) There exists $\bar{L} \in \mathbb{R}$ such that

$$L_k(x^{k+1}) \le \bar{L} \qquad \text{for all } k.$$

The next lemma ensures condition *(ii)* by comparison with a fixed feasible reference point.

Lemma 2 (Uniform AL bound from a feasible reference point) *Let x^{ref} be feasible for* (2)*. Suppose that $\rho_k \ge \rho_0 > 0$, $\|\widetilde{\Upsilon}^k\| \le M_{\mathrm{saf}}$, and the returned points satisfy* (23)*. Then*

$$L_k(x^{k+1}) \le L_k(x^{\mathrm{ref}}) \le f(w^{\mathrm{ref}}) + \frac{M_{\mathrm{saf}}^2}{2\rho_0} =: \bar{L}_{\mathrm{ref}} \qquad \text{for every } k. \tag{30}$$

Proof Feasibility gives $\Psi(x^{\mathrm{ref}}) \in C$, so

$$\operatorname{dist}\left(\Psi(x^{\mathrm{ref}}) + \frac{\bar{\Upsilon}^k}{\rho_k}, C\right) \le \frac{\|\bar{\Upsilon}^k\|}{\rho_k}.$$

Substitution into the squared-distance definition of L_k gives $L_k(x^{\mathrm{ref}}) \le f(w^{\mathrm{ref}}) + \|\bar{\Upsilon}^k\|^2/(2\rho_k)$. The initialization comparison and descent in (23) give $L_k(x^{k+1}) \le L_k(x^{\mathrm{ref}})$. The bounds on ρ_k and $\bar{\Upsilon}^k$ yield (30). □

Jia et al. [20, Theorem 4.3 and Corollary 4.4] established AM-stationarity of feasible accumulation points and M-stationarity under AM-regularity. In our algorithm, the two stationarity tests in Step 3 yield an overall stationarity error bounded by $\sqrt{2}\,\nu_{k+1}$. Applying their results, together with Proposition 4 and the regularity and stationarity equivalence results in Section 2, gives the following theorem.

Theorem 1 (AM- and M-stationarity of accumulation points) *Suppose that Algorithm 1 generates an infinite sequence $\{x^{k+1}\}$ and that $\nu_k \to 0$. Then the following statements hold:*

(i) Every feasible accumulation point of this sequence is AM-stationary for (2)*.*
(ii) Furthermore, if either condition (i) *or* (ii) *in Proposition 4 holds, then every accumulation point of this sequence is feasible and AM-stationary for* (2)*.*

For any feasible accumulation point $\bar{x} = (\bar{w}, \bar{u})$, if (2) *is AM-regular at $\bar{x}$, then $\bar{x}$ is M-stationary for* (2)*, and $\bar{w}$ is M-stationary for* (1)*.*

By Section 2.3, the AM-regularity required in the last assertion of Theorem 1 holds if (1) admits no singular M-multiplier at $\bar{w}$, in particular if active-set SDCMPCC-LICQ holds at $\bar{x}$.

Suppose that each Step 3 solve is completed and $\nu_k \to 0$. For a run continued with Step 1 omitted, the argument of [20, proof of Proposition 4.1(a) and discussion following Theorem 4.3], together with (27), gives $V_{k+1} \to 0$ if $\{\rho_k\}$ is bounded and $z^{k+1} \to_{\mathcal{K}} 0$ along every subsequence $x^{k+1} \to_{\mathcal{K}} \bar{x} \in \widetilde{\mathcal{F}}$. Since $V_{k+1} \le \|z^{k+1}\|$ and both stationarity residuals at x^{k+1} are at most ν_{k+1}, the test in Step 1 is satisfied after finitely many iterations for any positive stopping tolerances whenever $\{\rho_k\}$ is bounded or the iterates have a feasible accumulation point.

Corollary 1 (Local multiplier convergence) *Let $x^j = (w^j, u^j) \to \bar{x} = (\bar{w}, \bar{u}) \in \widetilde{\mathcal{F}}$, where active-set SDCMPCC-LICQ holds, and let $z^j, e^j \to 0$. If $t^j \in N_C(\Psi(x^j) - z^j) \times N_\Omega^{\lim}(u^j)$ satisfies $\mathcal{J}_{w^j} t^j = e^j - \nabla\varphi(x^j)$, then t^j converges to the unique M-stationarity multiplier tuple at $\bar{x}$.*

Proof The normal relations give $t^j \in \mathfrak{L}(w^j, u^j, z^j)$, so (19) yields $c\|t^j\| \le \|e^j\| + \|\nabla f(w^j)\|$ for all sufficiently large j. By continuity and closedness of the normal-cone graphs, every cluster point is an M-stationarity multiplier tuple at $\bar{x}$. Any two such tuples differ by an element of $\ker \mathcal{J}_{\bar{w}} \cap \mathfrak{L}(\bar{w}, \bar{u}, 0) = \{0\}$, proving convergence. □

Along any subsequence converging to a feasible point satisfying active-set SDCMPCC-LICQ, Corollary 1 applies to the updated multipliers and any Step 3 normal witness. Indeed, (25) and (26) give the required relations for $t^{k+1} = (\Upsilon^{k+1}, \Gamma^{k+1})$ with $e^{k+1} = \nabla L_k(x^{k+1}) + (0, \Gamma^{k+1})$ and $\|e^{k+1}\| \le \sqrt{2}\,\nu_{k+1}$.

Remark 3 (Comparison with Santos et al. [32]) Both methods retain semidefinite complementarity and penalize smooth linking constraints. Santos et al. impose approximate KKT conditions for the scalar representation of semidefinite complementarity. Their Theorem 6.1 gives AW-stationarity at feasible limits and AC-stationarity under the additional coupling $\delta_k \le \epsilon_k/(1+\gamma_k)$ in their Assumption 6.1. SPALM instead tests the limiting-normal distance directly in Step 3.

The next two sections give verifiable block conditions and procedures for generating slack candidates.

4 An inexact proximal alternating framework with verifiable block conditions

We solve the augmented Lagrangian subproblem at each outer iteration by proximal alternating updates. The analysis concerns accepted iterates satisfying the prescribed descent and stationarity tests.

4.1 Proximal block subproblems and the shifted projection formula

Throughout this section, consider the subproblem at the k-th outer iteration and write $(\rho, \Upsilon) := (\rho_k, \bar{\Upsilon}^k)$. These parameters remain unchanged during the inner solve. Set

$$L_\rho(w, u) := \mathcal{L}_\rho(w, u; \Upsilon),$$

where the augmented Lagrangian is given in (20). The AL subproblem is

$$\min_{w \in \mathcal{Z},\, u \in \Omega} L_\rho(w, u). \tag{31}$$

For fixed w, define the unregularized slack center

$$c_\rho(w) := \left(G(w) + \frac{M^G}{\rho}, H(w) + \frac{M^H}{\rho} \right) \in \mathcal{U}.$$

Up to a term independent of u, the slack-dependent part of $L_\rho(w, u)$ is $\frac{\rho}{2}\|u - c_\rho(w)\|^2$. Thus minimizing it over Ω is equivalent to projecting $c_\rho(w)$ onto Ω.

To stabilize both block updates and obtain sufficient decrease, we introduce proximal regularization. For proximal parameters $a_t > 0$ and $b_t > 0$, define

$$\Theta^t(w) := L_\rho(w, u^t) + \frac{a_t}{2} \left\| w - w^t \right\|^2 \tag{32}$$

and

$$\Phi^t(u) := L_\rho(w^{t+1}, u) + \frac{b_t}{2} \left\| u - u^t \right\|^2, \qquad u \in \Omega. \tag{33}$$

Completing the square shows that $\Phi^t(u)$ differs from $\frac{\rho+b_t}{2}\|u-\widehat{c}^t\|^2$ by a term independent of u, where

$$\widehat{c}^t := \frac{\rho c_\rho(w^{t+1}) + b_t u^t}{\rho + b_t}.$$

Hence an exact proximal slack update satisfies

$$u_\star^{t+1} \in \operatorname*{argmin}_{u\in\Omega} \|u-\widehat{c}^t\|^2. \tag{34}$$

Since Ω is nonempty and closed, (34) has a solution. Every exact projection gives $\Phi^t(u_\star^{t+1}) \le \Phi^t(u^t)$ and

$$-\nabla_u \Phi^t(u_\star^{t+1}) = (\rho+b_t)(\widehat{c}^t - u_\star^{t+1}) \in \widehat{N}_\Omega(u_\star^{t+1}) \subset N_\Omega^{\lim}(u_\star^{t+1}).$$

Lemma 3 (Lipschitz dependence of $\nabla_w L_\rho$ on slack variables) *Assume $f, g, h, G, H \in C^1$. For every bounded set $\mathcal{B} \subset \mathcal{Z}$, there exists $L_{wu} < \infty$ such that*

$$\|\nabla_w L_\rho(w,u) - \nabla_w L_\rho(w,u')\| \le L_{wu}\|u-u'\| \qquad \forall w \in \mathcal{B}, \quad \forall u,u' \in \mathcal{U}. \tag{35}$$

Here, for a linear operator $\mathcal{A}$ between finite-dimensional Euclidean spaces, $\|\mathcal{A}\|_{\mathrm{op}}$ denotes its induced operator norm. One may take

$$L_{wu} := \sqrt{2}\rho \sup_{w\in\mathcal{B}} \max\{\|G'(w)\|_{\mathrm{op}}, \|H'(w)\|_{\mathrm{op}}\}.$$

Proof Differentiating (20) and subtracting at $u=(Y,Z)$ and $u'=(Y',Z')$ gives

$$\begin{aligned}&\nabla_w L_\rho(w,u) - \nabla_w L_\rho(w,u') \\ &\quad = -\rho G'(w)^*(Y-Y') - \rho H'(w)^*(Z-Z').\end{aligned} \tag{36}$$

The triangle inequality, $\|Y-Y'\|_F + \|Z-Z'\|_F \le \sqrt{2}\|u-u'\|$, and boundedness of the continuous derivatives on $\overline{\mathcal{B}}$ prove the claim.

□

For later use, define the pointwise slack Lipschitz factor

$$L_{wu}(w) := \sqrt{2}\rho \max\left\{\|G'(w)\|_{\mathrm{op}}, \|H'(w)\|_{\mathrm{op}}\right\}.$$

Then, at the $(t+1)$-st accepted iterate, we write $L_{wu}^{t+1} := L_{wu}(w^{t+1})$.

For (85) and (89), $G(w) = X$ and $H(w) = -S$ act on separate components. The product norm in (36) therefore permits the smaller cross-Lipschitz constant ρ in the subsequent residual bounds, in place of the generic factor $L_{wu}(w)$ defined above.

4.2 Verifiable block updates and accepted iterates

A trial point becomes an accepted iterate only after both block tests below pass.

Algorithm 2 Inexact proximal alternating framework with verifiable block conditions

Input: An initial accepted point $x^0 = (w^0, u^0)$ with $u^0 \in \Omega$; positive proximal-parameter sequences $\{a_t\}$ and $\{b_t\}$; nonnegative tolerance sequences $\{\delta_w^t\}$ and $\{\delta_u^t\}$; a prescribed target $\nu > 0$.

Output: Accepted iterates $\{x^t\}$ satisfying the prescribed block conditions, together with normals $\Gamma^{t+1} \in N_\Omega^{\lim}(u^{t+1})$ certifying the slack residuals.

Set $t = 0$.

Step 1: inexact proximal w-update.

Invoke a w-block oracle to generate trial points for the proximal objective Θ^t in (32).

Accept a trial point as w^{t+1} only if

$$\Theta^t(w^{t+1}) \le \Theta^t(w^t) \tag{37}$$

and

$$\left\|\nabla_w L_\rho(w^{t+1}, u^t) + a_t(w^{t+1} - w^t)\right\| \le \delta_w^t. \tag{38}$$

Step 2: inexact proximal slack update.

Invoke a u-block oracle to generate feasible trial points for the proximal objective Φ^t in (33).

Accept a trial point as $u^{t+1} \in \Omega$ only if

$$\Phi^t(u^{t+1}) \le \Phi^t(u^t) \tag{39}$$

and

$$\operatorname{dist}\left(-\nabla_u \Phi^t(u^{t+1}), N_\Omega^{\lim}(u^{t+1})\right) \le \delta_u^t. \tag{40}$$

Retain a normal $\Gamma^{t+1} \in N_\Omega^{\lim}(u^{t+1})$ such that $\|\nabla_u \Phi^t(u^{t+1}) + \Gamma^{t+1}\| \le \delta_u^t$. The computable sufficient test (82) supplies such a normal.

Step 3: iterate update.

Set $x^{t+1} := (w^{t+1}, u^{t+1})$. If $\widehat{\mathcal{R}}_\rho^{t+1} \le \nu$, return this point and its normal. Otherwise increase t and repeat.

The slack residual condition (40) is equivalently

$$\operatorname{dist}\left(-\nabla_u L_\rho(w^{t+1}, u^{t+1}) - b_t(u^{t+1} - u^t), N_\Omega^{\lim}(u^{t+1})\right) \le \delta_u^t. \tag{41}$$

For example, L-BFGS with line search [26] can generate w-trials, and Section 5 supplies feasible slack trials. Budget exhaustion before acceptance is reported as *target not verified*.

For target $\nu > 0$, the full-residual bound (51) is checked after both updates. Returning x^{t+1} and its retained normal when $\widehat{\mathcal{R}}_\rho^{t+1} \le \nu$ implies Step 3 of Algorithm 1. Conversely, its two componentwise tests only imply a combined residual at most $\sqrt{2}\nu$. Omitting the stopping test defines the infinite accepted sequences below.

Both the existence argument and the convergence analysis use the initial sublevel set

$$\mathcal{S}_\rho(x^0) := \{x \in \mathcal{Z} \times \Omega : L_\rho(x) \le L_\rho(x^0)\}. \tag{42}$$

If this set is bounded, it is compact, since L_ρ is continuous and $\mathcal{Z} \times \Omega$ is closed. In particular, L_ρ attains a finite minimum on it.

Proposition 5 (Existence of accepted block updates) *Assume that $f, g, h, G, H \in C^1$ and that $\mathcal{S}_\rho(x^0)$ is bounded. At every accepted iterate x^t, and for any $a_t, b_t > 0$, exact minimizers*

$$w_\star^{t+1} \in \arg\min_{w \in \mathcal{Z}} \Theta^t(w), \qquad u_\star^{t+1} \in \arg\min_{u \in \Omega} \Phi^t(u)$$

exist, where $w^{t+1} = w_\star^{t+1}$ is used to define Φ^t. These minimizers satisfy the two block acceptance conditions with zero residuals.

Proof The accepted descent conditions give $L_\rho(x^t) \le L_\rho(x^0)$. The set $\{w : \Theta^t(w) \le \Theta^t(w^t)\}$ is nonempty and closed. Every point in this set satisfies

$$L_\rho(w, u^t) \le \Theta^t(w) \le L_\rho(x^t) \le L_\rho(x^0),$$

so the set is bounded by the assumption on $\mathcal{S}_\rho(x^0)$. Thus Θ^t attains a global minimum, and its gradient vanishes there. This gives (37) and (38) with zero residual. The projection representation (34) provides an exact slack update satisfying (39) and (40) with zero residual. Induction therefore gives accepted updates at every iteration when exact block minimizers are available. □

4.3 Convergence and finite-accuracy guarantees for accepted sequences

We establish convergence and finite-accuracy guarantees using sufficient decrease and estimates of the full stationarity residual.

Lemma 4 (Sufficient decrease and vanishing steps) *Suppose that Algorithm 2 generates a finite or infinite accepted sequence $\{x^t = (w^t, u^t)\}$, with a constant $\underline{c} > 0$ such that*

$$a_t \ge \underline{c}, \qquad b_t \ge \underline{c} \qquad \forall t.$$

Then every accepted transition satisfies

$$L_\rho(w^t, u^t) - L_\rho(w^{t+1}, u^t) \ge \frac{a_t}{2} \left\|w^{t+1} - w^t\right\|^2, \tag{43}$$

and

$$L_\rho(w^{t+1}, u^t) - L_\rho(w^{t+1}, u^{t+1}) \ge \frac{b_t}{2} \left\|u^{t+1} - u^t\right\|^2. \tag{44}$$

Consequently,

$$L_\rho(x^t) - L_\rho(x^{t+1}) \ge \frac{\underline{c}}{2} \left\|x^{t+1} - x^t\right\|^2. \tag{45}$$

If $\mathcal{S}_\rho(x^0)$ is bounded and the sequence is infinite, then

$$\sum_{t=0}^{\infty} \left\|x^{t+1} - x^t\right\|^2 < \infty,$$

and hence

$$\left\|x^{t+1} - x^t\right\| \to 0. \tag{46}$$

In particular,

$$\left\|w^{t+1} - w^t\right\| \to 0, \qquad \left\|u^{t+1} - u^t\right\| \to 0.$$

Proof Expanding (37) and (39) gives (43) and (44); adding them proves (45). All accepted points remain in $\mathcal{S}_\rho(x^0)$. If this set is bounded, continuity gives a finite lower bound m_ρ for L_ρ there. Summation yields

$$\frac{\underline{c}}{2}\sum_{t=0}^{T}\|x^{t+1}-x^t\|^2 \le L_\rho(x^0)-L_\rho(x^{T+1}) \le L_\rho(x^0)-m_\rho,$$

which proves square summability and (46).

□

For $x=(w,u)\in\mathcal{Z}\times\Omega$, define the full M-stationarity residual of the inner subproblem by

$$\mathcal{R}_\rho(x) := \Big(\|\nabla_w L_\rho(w,u)\|^2 + \operatorname{dist}^2\big(-\nabla_u L_\rho(w,u), N_\Omega^{\lim}(u)\big)\Big)^{1/2}. \tag{47}$$

Lemma 5 (One-step residual bound and normal-certificate transfer) *Assume that $f,g,h,G,H\in C^1$. Let $x^t=(w^t,u^t)$ and $x^{t+1}=(w^{t+1},u^{t+1})$ be consecutive accepted iterates of Algorithm 2, with $a_t,b_t>0$. Write*

$$\Delta w^t := w^{t+1}-w^t, \qquad \Delta u^t := u^{t+1}-u^t = (\Delta Y^t, \Delta Z^t).$$

There exists $\Gamma^{t+1}\in N_\Omega^{\lim}(u^{t+1})$ such that the block errors

$$e_w^t := \nabla_w L_\rho(w^{t+1},u^t)+a_t\Delta w^t, \qquad e_u^t := \nabla_u \Phi^t(u^{t+1})+\Gamma^{t+1}$$

satisfy $\|e_w^t\|\le\delta_w^t$ and $\|e_u^t\|\le\delta_u^t$. For any such normal,

$$\begin{aligned}\nabla_w L_\rho(x^{t+1}) &= e_w^t - a_t\Delta w^t - \rho G'(w^{t+1})^*\Delta Y^t - \rho H'(w^{t+1})^*\Delta Z^t,\\ \nabla_u L_\rho(x^{t+1})+\Gamma^{t+1} &= e_u^t - b_t\Delta u^t.\end{aligned} \tag{48}$$

Consequently,

$$\|\nabla_w L_\rho(x^{t+1})\| \le L_{wu}^{t+1}\|\Delta u^t\| + a_t\|\Delta w^t\| + \delta_w^t \tag{49}$$

and

$$\begin{aligned}\operatorname{dist}\big(-\nabla_u L_\rho(x^{t+1}), N_\Omega^{\lim}(u^{t+1})\big) &\le \|\nabla_u L_\rho(x^{t+1})+\Gamma^{t+1}\|\\ &\le b_t\|\Delta u^t\|+\delta_u^t.\end{aligned} \tag{50}$$

Define the computable upper bound

$$\widehat{\mathcal{R}}_\rho^{t+1} := \Big[\big(L_{wu}^{t+1}\|\Delta u^t\|+a_t\|\Delta w^t\|+\delta_w^t\big)^2 + \big(b_t\|\Delta u^t\|+\delta_u^t\big)^2\Big]^{1/2} \tag{51}$$

and the full error vector

$$E^{t+1} := \big(\nabla_w L_\rho(x^{t+1}), \nabla_u L_\rho(x^{t+1})+\Gamma^{t+1}\big).$$

Then

$$\mathcal{R}_\rho(x^{t+1}) \le \|E^{t+1}\| \le \widehat{\mathcal{R}}_\rho^{t+1}. \tag{52}$$

Proof The distance to the nonempty closed set $N_\Omega^{\lim}(u^{t+1})$ is attained; a normal retained from (82) also suffices. The definitions of e_w^t, e_u^t and (36) give (48). Taking norms and applying Lemma 3 proves the bounds. The same normal is used throughout; the proximal term $-b_t\Delta u^t$ is part of the error.

□

Theorem 2 (Residual convergence and M-stationarity) *Assume that $f,g,h,G,H\in C^1$, and that there exist constants $0<\underline{c}\le\overline{c}<\infty$ such that*

$$\underline{c}\le a_t, b_t\le\overline{c} \qquad \text{for all } t.$$

Suppose that Algorithm 2 generates an infinite accepted sequence $\{x^t=(w^t,u^t)\}$, that $\mathcal{S}_\rho(x^0)$ is bounded, and that $\delta_w^t,\delta_u^t\to 0$. Then:

(i) The accepted sequence $\{x^t\}$ is bounded.

(ii) The computable upper bound and the full residual satisfy

$$\widehat{\mathcal{R}}_\rho^{t+1}\to 0, \qquad \mathcal{R}_\rho(x^{t+1})\to 0. \tag{53}$$

(iii) Every accumulation point $\bar{x}=(\bar{w},\bar{u})$ of $\{x^t\}$ is M-stationary for the AL subproblem (31); that is,

$$\nabla_w L_\rho(\bar{w},\bar{u})=0, \qquad -\nabla_u L_\rho(\bar{w},\bar{u})\in N_\Omega^{\lim}(\bar{u}).$$

Proof Lemma 4 places all accepted iterates in the compact set $\mathcal{S}_\rho(x^0)$ and gives $\|x^{t+1} - x^t\| \to 0$. This proves (i). Continuity of G' and H' implies boundedness of $\{L_{wu}^{t+1}\}$. Together with the parameter bounds and vanishing block errors, (51) gives $\widehat{\mathcal{R}}_\rho^{t+1} \to 0$. Lemma 5 then proves (ii).

Let $x^{t_j} \to \bar{x}$. Vanishing steps give $x^{t_j+1} \to \bar{x}$, and closedness of Ω gives $\bar{u} \in \Omega$. Use the normals Γ^{t_j+1} from Lemma 5. Since $\|E^{t_j+1}\| \to 0$, continuity of the two gradients yields

$$\nabla_w L_\rho(\bar{x}) = 0, \qquad \Gamma^{t_j+1} \to -\nabla_u L_\rho(\bar{x}).$$

The closed graph of the limiting normal-cone mapping implies $-\nabla_u L_\rho(\bar{x}) \in N_\Omega^{\lim}(\bar{u})$, proving (iii). □

Propositions 10 and 12 verify the sublevel-set boundedness assumption for our numerical models.

Corollary 2 (Finite-accuracy output) *Assume C^1 data, boundedness of $\mathcal{S}_\rho(x^0)$, $0 < \underline{c} \le a_t, b_t \le \bar{c} < \infty$, and an infinite accepted sequence. Fix $\nu > 0$. If either* (i) $\delta_w^t, \delta_u^t \to 0$, *or* (ii) $0 \le \delta_w^t \le c_w\nu$, $0 \le \delta_u^t \le c_u\nu$ *with* $\kappa := \sqrt{c_w^2 + c_u^2} < 1$, *then eventually*

$$\mathcal{R}_\rho(x^{t+1}) \le \|E^{t+1}\| \le \widehat{\mathcal{R}}_\rho^{t+1} < \nu.$$

For the AL subproblem at outer iteration k with parameters $(\rho_k, \bar{Y}^k)$, choose $x^{k,0} \in \mathcal{Z} \times \Omega$, and impose the preceding assumptions on its initial sublevel set and accepted sequence. Denote its accepted iterates, retained normals and bounds by $x^{k,t}$, $\Gamma^{k,t+1}$ and $\widehat{\mathcal{R}}_k^{t+1}$. The initial point need not equal x^k, and the sublevel-set bound need not be uniform in k. If a feasible reference point is supplied, choose $x^{k,0}$ by (23). *Lemma 4 then gives $L_k(x^{k,t}) \le L_k(x^{k,0})$ for every accepted iterate, so the returned point also satisfies the value comparison. Under either regime with target ν_{k+1}, the index*

$$t_k := \min\{t \in \mathbb{N}_0 : \widehat{\mathcal{R}}_k^{t+1} \le \nu_{k+1}\} \tag{54}$$

is finite. Returning $x^{k+1} = x^{k,t_k+1}$ and $\Gamma^{k+1} = \Gamma^{k,t_k+1} \in N_\Omega^{\lim}(u^{k+1})$ gives

$$\|\big(\nabla_w L_k(x^{k+1}), \nabla_u L_k(x^{k+1}) + \Gamma^{k+1}\big)\| \le \nu_{k+1},$$

and hence both conditions in Step 3 of Algorithm 1.

Proof In case *(i)*, apply Theorem 2. In case *(ii)*, Lemma 4 gives vanishing steps, while L_{wu}^{t+1}, a_t, b_t are bounded. Thus (51) gives $\limsup_t \widehat{\mathcal{R}}_\rho^{t+1} \le \kappa\nu < \nu$. The returned bound follows from Lemma 5 at that same point and normal. □

For example, $c_w = c_u = 1/2$ is valid. The retained normal can also be used in (29) when $\nu_{k+1} \le \varepsilon_{\text{stat}}$, or when its slack residual is directly verified against $\varepsilon_{\text{stat}}$.

Corollary 3 (Explicit residual bounds in accepted iterations) *Assume that $f, g, h, G, H \in C^1$, that $\mathcal{S}_\rho(x^0)$ is bounded, and that $0 < \underline{c} \le a_t, b_t \le \bar{c} < \infty$ for all accepted transitions. Let $\mathcal{B}_\rho$ be the projection of $\mathcal{S}_\rho(x^0)$ onto the w-space, and define*

$$m_\rho := \min_{x \in \mathcal{S}_\rho(x^0)} L_\rho(x), \quad \Delta_\rho := L_\rho(x^0) - m_\rho,$$

$$\overline{L}_{wu} := \sup_{w \in \mathcal{B}_\rho} L_{wu}(w), \quad K_\rho := \sqrt{2\bar{c}^{\,2} + \overline{L}_{wu}^{\,2}}.$$

These constants are finite. For any $N \ge 1$ accepted transitions, put

$$\epsilon_t := \sqrt{(\delta_w^t)^2 + (\delta_u^t)^2}, \qquad E_N := \sum_{t=0}^{N-1} \epsilon_t^2.$$

Then

$$\min_{0 \le t < N} \big(\widehat{\mathcal{R}}_\rho^{t+1}\big)^2 \le \frac{4K_\rho^2 \Delta_\rho/\underline{c} + 2E_N}{N}. \tag{55}$$

In particular, along an infinite accepted sequence with $E_\infty := \sum_{t=0}^{\infty} \epsilon_t^2 < \infty$,

$$\min_{0 \le t < N} \widehat{\mathcal{R}}_\rho^{t+1} \le \sqrt{\frac{4K_\rho^2 \Delta_\rho/\underline{c} + 2E_\infty}{N}} = O(N^{-1/2}).$$

If instead $\epsilon_t \le \kappa\nu$ for a fixed $\nu > 0$ and $0 \le \kappa < 1$, then

$$\min_{0 \le t < N} \widehat{\mathcal{R}}_\rho^{t+1} \le \kappa\nu + K_\rho \sqrt{\frac{2\Delta_\rho}{\underline{c}N}}. \tag{56}$$

The same upper bounds hold with $\mathcal{R}_\rho(x^{t+1})$ in place of $\widehat{\mathcal{R}}_\rho^{t+1}$.

Proof Compactness of $\mathcal{S}_\rho(x^0)$ and continuity of the data and their derivatives ensure finiteness of the constants. Let $s_t := \|x^{t+1} - x^t\|$. By the product-norm triangle inequality in (51) and

$$\left\| \begin{pmatrix} a_t & L_{wu}^{t+1} \\ 0 & b_t \end{pmatrix} \right\|_{\mathrm{op}} \le \left\| \begin{pmatrix} a_t & L_{wu}^{t+1} \\ 0 & b_t \end{pmatrix} \right\|_F \le K_\rho,$$

we obtain $\widehat{\mathcal{R}}_\rho^{t+1} \le K_\rho s_t + \epsilon_t$. Summing the finite-transition decrease inequality (45) gives

$$\sum_{t=0}^{N-1} s_t^2 \le \frac{2\big(L_\rho(x^0) - L_\rho(x^N)\big)}{\underline{c}} \le \frac{2\Delta_\rho}{\underline{c}}.$$

Squaring the pointwise residual bound, using $(p+q)^2 \le 2p^2 + 2q^2$, and averaging over $t = 0, \dots, N-1$ proves (55). The square-summable case follows by replacing E_N with E_∞. If $\epsilon_t \le \kappa\nu$, choose an index minimizing s_t to obtain

$$\min_{0\le t<N} \widehat{\mathcal{R}}_\rho^{t+1} \le \kappa\nu + K_\rho \min_{0\le t<N} s_t \le \kappa\nu + K_\rho \sqrt{\frac{2\Delta_\rho}{\underline{c}N}}.$$

Finally, Lemma 5 gives the bounds for $\mathcal{R}_\rho$. □

Under the fixed-accuracy allocation in Corollary 2, (56) guarantees a point with $\widehat{\mathcal{R}}_\rho^{t+1} < \nu$ among the first N accepted transitions whenever

$$N > B_\nu := \frac{2K_\rho^2 \Delta_\rho}{\underline{c}(1-\kappa)^2\nu^2}.$$

Thus $\lfloor B_\nu \rfloor + 1$ is a sufficient number of accepted transitions for the first tolerance hit.

A known lower bound for L_ρ on $\mathcal{S}_\rho(x^0)$ may replace m_ρ in these estimates. Usable bounds for the remaining constants then give a numerical upper bound on the number of accepted transitions required to reach the target.

For a fixed AL subproblem and initial sublevel set, the bound is $O(\nu^{-2})$ when $\underline{c} > 0$, $\bar{c}$, and $\kappa < 1$ are independent of ν. It counts accepted transitions; the cost of generating and verifying trial points is additional, and the constants may vary across outer iterations.

In the next section, we use the shifted projection representation to generate feasible slack candidates, which are then checked against the descent and stationarity tests.

5 The slack u-subproblem: feasible-candidate generation

At the t-th inner iteration, write

$$\widehat{c}^t = (A^t, B^t), \qquad A^t, B^t \in \mathbb{S}^n.$$

Since the projection routines developed below apply to a generic center, we suppress the iteration index and write $(A, B) = (A^t, B^t)$. The exact proximal slack update then reduces to

$$\min_{(Y,Z)\in\Omega} \ \|Y - A\|_F^2 + \|Z - B\|_F^2 . \tag{57}$$

Remark 4 (Complexity of exact projection) When n is part of the input, deciding whether the optimal value of (57) is at most a given rational threshold is NP-hard for rational A, B. A reduction from Simple Max-Cut [13] constructs centers $A = D + \varepsilon C$ and $B = -D + \varepsilon C$, with D positive diagonal and C encoding the graph. A uniform second-order estimate allows the maximum cut size to be recovered from the projection value. The rational construction is computable in polynomial time, yielding the claimed reduction. The complete proof is given in Appendix E.

An exact solution of (57) is computationally expensive. We therefore use a common diagonal reconstruction to generate feasible candidates through either a Huber–Stiefel search or a fixed eigenbasis of A or B.

For the proximal slack update, set $q_t(u) := \|u - \widehat{c}^t\|^2$. Completing the square in (33) gives

$$\Phi^t(u) = C_t + \frac{\rho + b_t}{2} q_t(u), \qquad \Phi^t(u_{\mathrm{cand}}) \le \Phi^t(u^t) \iff q_t(u_{\mathrm{cand}}) \le q_t(u^t), \tag{58}$$

where C_t is independent of u. Thus candidate generation is followed by a comparison of the true squared distances and a separate normal-residual test at the same candidate.

5.1 Diagonal decoupling for a fixed orthogonal matrix

Every $(Y, Z) \in \Omega$ has a joint representation $Y = P\,\mathrm{Diag}(y)P^\top$, $Z = P\,\mathrm{Diag}(z)P^\top$, with $P \in \mathbb{O}^n$, $y, z \in \mathbb{R}^n$, $y \geq 0$, $z \leq 0$ and $y_i z_i = 0$; conversely, these conditions imply membership in Ω.

For such a matrix P, define the diagonal vectors

$$d_A(P) := \mathrm{diag}(P^\top A P), \qquad d_B(P) := \mathrm{diag}(P^\top B P),$$

whose i-th components are denoted by $d_{A,i}(P)$ and $d_{B,i}(P)$, respectively.

For fixed P, orthogonal invariance of the Frobenius norm implies that the off-diagonal terms are independent of y and z. Hence, the minimization over y and z separates into the following scalar complementarity problems, one for each $i = 1, \ldots, n$, as in [15, Section 4, equations (27)–(29)]:

$$\min_{y_i, z_i} (y_i - d_{A,i}(P))^2 + (z_i - d_{B,i}(P))^2 \quad \text{s.t.} \quad y_i \geq 0,\ z_i \leq 0,\ y_i z_i = 0.$$

The minimizers over the two complementarity branches are

$$(y_i, z_i) = ((d_{A,i}(P))_+, 0) \quad \text{and} \quad (y_i, z_i) = (0, -(d_{B,i}(P))_-),$$

respectively. Select the better branch by

$$\chi_i(P) := \begin{cases} 1, & (d_{A,i}(P))_+^2 \geq (d_{B,i}(P))_-^2, \\ 0, & (d_{A,i}(P))_+^2 < (d_{B,i}(P))_-^2. \end{cases}$$

Ties are resolved by choosing $\chi_i(P) = 1$. Define the scalar reconstruction by

$$y_i(P) := \chi_i(P)(d_{A,i}(P))_+, \qquad z_i(P) := -(1 - \chi_i(P))(d_{B,i}(P))_-.$$

Collect these coefficients into the vectors

$$y(P) := \big(y_1(P), \ldots, y_n(P)\big)^\top, \qquad z(P) := \big(z_1(P), \ldots, z_n(P)\big)^\top.$$

The corresponding matrices are

$$Y(P) := P\,\mathrm{Diag}(y(P))P^\top, \qquad Z(P) := P\,\mathrm{Diag}(z(P))P^\top. \tag{59}$$

Thus $u(P) := (Y(P), Z(P)) \in \Omega$.

For each fixed P, the reconstructed candidate minimizes the diagonal subproblem. Its objective value is

$$\begin{aligned} V(P) &:= \|Y(P) - A\|_F^2 + \|Z(P) - B\|_F^2 \\ &= \|A\|_F^2 + \|B\|_F^2 - F(P), \end{aligned} \tag{60}$$

where

$$\begin{aligned} F(P) &:= \sum_{i=1}^n \max\{(d_{A,i}(P))_+^2, (d_{B,i}(P))_-^2\} \\ &= \sum_{i=1}^n \big(\max\{d_{A,i}(P), -d_{B,i}(P), 0\}\big)^2. \end{aligned}$$

Thus, the search over P reduces to the nonsmooth maximization problem $\max_{P \in \mathbb{O}^n} F(P)$.

5.2 Huber–Stiefel feasible-candidate generation

To obtain a continuously differentiable approximation of F, we smooth the maximum of $d_{A,i}(P)$ and $-d_{B,i}(P)$ in each summand. For a smoothing parameter $\delta > 0$, we use the following Huber-type approximation of the absolute value [17]:

$$H_\delta(s) := \begin{cases} \dfrac{s^2}{2\delta} + \dfrac{\delta}{2}, & |s| \le \delta, \\ |s|, & |s| > \delta, \end{cases}$$

with derivative

$$H'_\delta(s) := \begin{cases} s/\delta, & |s| \le \delta, \\ \operatorname{sign}(s), & |s| > \delta. \end{cases}$$

A continuously differentiable approximation of $\max\{a, b\}$ is

$$M_\delta(a, b) := \frac{a + b + H_\delta(a - b)}{2}.$$

Since $0 \le H_\delta(s) - |s| \le \frac{\delta}{2}$, the smoothed maximum satisfies

$$0 \le M_\delta(a, b) - \max\{a, b\} \le \frac{\delta}{4}. \tag{61}$$

Thus $M_\delta(a, b)$ converges uniformly to $\max\{a, b\}$ as $\delta \downarrow 0$.

Set

$$m_i(P) := M_\delta(d_{A,i}(P), -d_{B,i}(P)) = \frac{d_{A,i}(P) - d_{B,i}(P) + H_\delta(d_{A,i}(P) + d_{B,i}(P))}{2}.$$

The smoothed Stiefel objective is

$$F_\delta(P) := \sum_{i=1}^{n} \big(m_i(P)\big)_+^2, \qquad P \in \mathbb{O}^n. \tag{62}$$

Since the scalar function $s \mapsto (s_+)^2$ is continuously differentiable and M_δ is C^1, the function F_δ is continuously differentiable on $\mathbb{O}^n$.

The following result quantifies the approximation error uniformly over $\mathbb{O}^n$.

Proposition 6 (Uniform approximation of the Stiefel objective) *Let*

$$R_0 := \max\{\|A\|_2, \|B\|_2\}.$$

Then, for every $\delta > 0$ and every $P \in \mathbb{O}^n$,

$$0 \le F_\delta(P) - F(P) \le n\left(\frac{R_0\delta}{2} + \frac{\delta^2}{16}\right). \tag{63}$$

Hence $F_\delta \to F$ uniformly on $\mathbb{O}^n$ as $\delta \downarrow 0$.

Proof For $P \in \mathbb{O}^n$, set $q_i(P) := \max\{d_{A,i}(P), -d_{B,i}(P)\}$. By (61), $0 \le m_i(P) - q_i(P) \le \delta/4$. Since the mapping $s \mapsto s_+$ is nondecreasing and 1-Lipschitz,

$$0 \le (m_i(P))_+ - (q_i(P))_+ \le \delta/4. \tag{64}$$

Since $0 \le (q_i(P))_+ \le R_0$, the difference of squares is at most $(\delta/4)(2R_0 + \delta/4)$. Summing over i proves (63). □

To carry out the Stiefel search, we compute the Euclidean gradient of F_δ. Let

$$\zeta_i := H'_\delta(d_{A,i}(P) + d_{B,i}(P)), \qquad \gamma_i := 2(m_i(P))_+,$$

and define

$$D_\gamma := \operatorname{Diag}(\gamma_1, \ldots, \gamma_n), \qquad D_{\gamma\zeta} := \operatorname{Diag}(\gamma_1\zeta_1, \ldots, \gamma_n\zeta_n).$$

By direct differentiation and the chain rule, the Euclidean gradient of F_δ is

$$\nabla_P F_\delta(P) = (A - B)PD_\gamma + (A + B)PD_{\gamma\zeta}. \tag{65}$$

Under the Frobenius metric, projection onto $T_P\mathbb{O}^n = \{\eta : P^\top\eta + \eta^\top P = 0\}$ gives the Riemannian gradient [10,1]:

$$\operatorname{grad} F_\delta(P) = \nabla_P F_\delta(P) - P\operatorname{sym}\big(P^\top \nabla_P F_\delta(P)\big). \tag{66}$$

Here $\operatorname{sym}(C) := (C + C^\top)/2$ for any square matrix C.

For the manifold search below, we use the standard QR-based retraction [1],

$$R_P(\eta) := \operatorname{qf}(P + \eta),$$

where $\operatorname{qf}(X)$ denotes the orthogonal factor Q in the QR factorization $X = QR$, with the diagonal entries of R chosen to be positive.

Since F_δ is continuously differentiable and R_P is a retraction, whenever $\operatorname{grad} F_\delta(P) \neq 0$, the ascent direction $\eta = \operatorname{grad} F_\delta(P)$ satisfies the standard ascent Armijo condition for all sufficiently small $\tau > 0$. Hence, the backtracking procedure in the algorithm below terminates after finitely many reductions.

Algorithm 3 Huber–Stiefel feasible-candidate generator

Input: $A, B \in \mathbb{S}^n$; initial $P^0 \in \mathbb{O}^n$; smoothing parameter $\delta > 0$; Armijo parameters $\alpha > 0$, $\beta_{\mathrm{ls}} \in (0,1)$, $\sigma \in (0,1)$; stopping tolerances $\eta_{\mathrm{st}} > 0$, $\eta_F > 0$; maximum number of iterations $K_{\max}$.

Output: A feasible candidate $u_{\mathrm{cand}} = (Y_{\mathrm{cand}}, Z_{\mathrm{cand}}) \in \Omega$.

Set $s = 0$.

Step 1: Gradient and ascent direction.

Compute $\operatorname{grad} F_\delta(P^s)$ from (65)–(66) and set $\eta^s := \operatorname{grad} F_\delta(P^s)$.

Step 2: Armijo backtracking and update.

Starting from $\tau = \alpha$, reduce $\tau \leftarrow \beta_{\mathrm{ls}}\tau$ until, with

$$P_{\mathrm{cand}} := R_{P^s}(\tau\eta^s),$$

the Armijo condition

$$F_\delta(P_{\mathrm{cand}}) - F_\delta(P^s) \geq \sigma\tau\,\langle \operatorname{grad} F_\delta(P^s), \eta^s\rangle \tag{67}$$

holds. Set $P^{s+1} := P_{\mathrm{cand}}$.

Step 3: Stiefel stopping tests.

If

$$\|\operatorname{grad} F_\delta(P^{s+1})\|_F < \eta_{\mathrm{st}}, \qquad \text{or} \qquad |F_\delta(P^{s+1}) - F_\delta(P^s)| < \eta_F,$$

or if $s + 1 \geq K_{\max}$, proceed to Step 4. Otherwise, set $s \leftarrow s + 1$ and return to Step 1.

Step 4: Feasible reconstruction.

Set

$$u_{\mathrm{cand}} := \big(Y(P^{s+1}), Z(P^{s+1})\big)$$

using (59), and return $u_{\mathrm{cand}} \in \Omega$.

Post-verification and acceptance. By (60), $q_t(u(P)) = V(P)$. If P^0 is a common eigenbasis of the current feasible point u^t, scalar minimization yields $V(P^0) \leq q_t(u^t)$. Increasing F_δ, however, need not preserve this bound.

Define

$$E_\delta := n\left(\frac{R_0\delta}{2} + \frac{\delta^2}{16}\right).$$

By (63), for any $P_0, P_1 \in \mathbb{O}^n$,

$$F(P_1) - F(P_0) \geq F_\delta(P_1) - F_\delta(P_0) - E_\delta. \tag{68}$$

Thus, under the above initialization, a smoothed gain of at least E_δ from P^0 ensures $q_t(u(P_1)) = V(P_1) \leq V(P^0) \leq q_t(u^t)$.

The gain condition is sufficient but not necessary for descent. Acceptance is determined by the true-value comparison in (58) and the certificate test (82). The following example explains why the normal-residual test is needed in addition to the Stiefel stopping criteria.

Remark 5 (Stiefel stopping versus slack verification) A zero Stiefel residual need not give a small slack residual. Take

$$A = \operatorname{Diag}(2,0), \quad B = 0, \quad P = 2^{-1/2}\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad 0 < \delta < 1.$$

Then $d_A(P) = (1,1)^\top$, $d_B(P) = 0$, $m_i = 1$, $\gamma_i = 2$ and $\zeta_i = 1$. Thus $\nabla_P F_\delta(P) = 4AP$ and $\operatorname{grad} F_\delta(P) = 0$, but $u(P) = (I, 0)$ and

$$\operatorname{dist}\big((\rho + b_t)(A - I, B), N_\Omega^{\lim}(I,0)\big) = (\rho + b_t)\|A - I\|_F = (\rho + b_t)\sqrt{2}.$$

Here $N_\Omega^{\lim}(I,0) = \{0\} \times \mathbb{S}^2$, so the Stiefel stopping tests cannot replace the normal-residual test.

5.3 Fixed-basis feasible-candidate generation

Feasible candidates can also be generated using a fixed eigenbasis of either A or B. We describe this construction below and derive computable objective-gap bounds in the next subsection. Let $A = P_A \Lambda_A P_A^\top$ and $B = P_B \Lambda_B P_B^\top$ be orthogonal eigendecompositions, where $\Lambda_A = \mathrm{Diag}(\lambda(A)), \Lambda_B = \mathrm{Diag}(\lambda(B))$. For each $i = 1, \ldots, n$, the scalar branch rule in Subsection 5.1 gives

$$\begin{aligned} y_i(P_A) &= \chi_i(P_A)(\lambda_i(A))_+, & z_i(P_A) &= -(1-\chi_i(P_A))\big((P_A^\top B P_A)_{ii}\big)_-, \\ y_i(P_B) &= \chi_i(P_B)\big((P_B^\top A P_B)_{ii}\big)_+, & z_i(P_B) &= -(1-\chi_i(P_B))(\lambda_i(B))_-. \end{aligned}$$

Reconstructing the matrices in their respective bases according to (59), we obtain

$$Y_A := P_A \,\mathrm{Diag}(y(P_A)) P_A^\top, \qquad Z_A := P_A \,\mathrm{Diag}(z(P_A)) P_A^\top, \qquad V_A := V(P_A), \tag{69}$$

$$Y_B := P_B \,\mathrm{Diag}(y(P_B)) P_B^\top, \qquad Z_B := P_B \,\mathrm{Diag}(z(P_B)) P_B^\top, \qquad V_B := V(P_B). \tag{70}$$

Both candidates lie in Ω, so V_A and V_B are upper bounds on the optimal value of the slack projection problem. They may be tried before or alongside the Stiefel search, subject to the same descent and normal-residual tests.

5.4 Objective-gap bounds for fixed-basis candidates

Let

$$v_\star := \min_{(Y,Z)\in\Omega} \left\{ \|Y - A\|_F^2 + \|Z - B\|_F^2 \right\} \tag{71}$$

denote the optimal value of (57).

Keeping the eigenpair ordering of A fixed, define

$$U_{\mathrm{spec}} := \max_{\pi} \sum_{i=1}^{n} \max\left\{ (\lambda_i(A))_+^2, (\lambda_{\pi(i)}(B))_-^2 \right\}, \tag{72}$$

where π ranges over all permutations of $\{1, \ldots, n\}$. Equivalently, U_{spec} is the sum of the n largest of the $2n$ values $(\lambda_i(A))_+^2$ and $(\lambda_i(B))_-^2$, $i = 1, \ldots, n$. Set

$$L_{\mathrm{spec}} := \|A\|_F^2 + \|B\|_F^2 - U_{\mathrm{spec}}. \tag{73}$$

To attain the maximum in (72), pair the values $(\lambda_i(A))_+^2$ and $(\lambda_j(B))_-^2$ in opposite sorted orders. Keep the eigenpair ordering of A fixed and reorder the eigenvalues of B and the corresponding columns of P_B accordingly. Relabelling the reordered eigenvalues gives

$$U_{\mathrm{spec}} = \sum_{i=1}^{n} \max\left\{ (\lambda_i(A))_+^2, (\lambda_i(B))_-^2 \right\}. \tag{74}$$

Using the resulting eigenvector matrices, define

$$T := (P_A^\top P_B) \circ (P_A^\top P_B),$$

where $\circ$ denotes the Hadamard product. The following bound relates the fixed-A objective gap to the alignment of the paired eigenbases.

Proposition 7 (Objective-gap bound for the fixed-A candidate) *Under the pairing* (74)*, with* $R_0 = \max\{\|A\|_2, \|B\|_2\}$*, we have*

$$0 \le V_A - v_\star \le V_A - L_{\mathrm{spec}} \le 2\sqrt{n}\, R_0 \, \|I - T\|_2 \, \|B\|_F. \tag{75}$$

The proof, including the pairing argument, is given in Appendix B.

By symmetry, the fixed-B candidate satisfies

$$0 \le V_B - v_\star \le V_B - L_{\mathrm{spec}} \le 2\sqrt{n}\, R_0 \, \|I - T\|_2 \, \|A\|_F,$$

since $\|I - T^\top\|_2 = \|I - T\|_2$.

The factor $\|I-T\|_2$ depends on the chosen eigenbases and the pairing (74), including choices within repeated eigenspaces. $AB = BA$ guarantees neither $T = I$ under this pairing nor $v_\star = L_{\mathrm{spec}}$. Under the same pairing condition, $T = I$ implies $V_A = V_B = v_\star = L_{\mathrm{spec}}$.

Thus, with $V_{\min} := \min\{V_A, V_B\}$,

$$0 \le V_{\min} - v_\star \le V_{\min} - L_{\mathrm{spec}}.$$

$V_{\min} - L_{\mathrm{spec}}$ is a computable upper bound on the objective gap of the better fixed-basis candidate. A zero value certifies global optimality, whereas a positive value leaves optimality undecided. Acceptance is evaluated separately through the descent and normal-residual tests.

5.5 Limiting-normal certificates and slack verification

We now turn to the normal-residual condition required for slack acceptance. For $u = (Y, Z) \in \Omega$ and a test vector $\xi = (\xi^Y, \xi^Z) \in \mathcal{U}$, we construct computable bounds on $\mathrm{dist}(\xi, N_\Omega^{\lim}(u))$. Throughout this subsection, $u \in \Omega$ and α, β, γ are the exact positive, zero, and negative spectral index sets of $Y + Z$. We use the joint spectral representation from Subsection 2.1 to express the test vector in spectral coordinates:

$$\widetilde{\xi}^Y := P^\top \xi^Y P, \qquad \widetilde{\xi}^Z := P^\top \xi^Z P, \qquad A_\beta := \widetilde{\xi}^Y_{\beta\beta}, \quad B_\beta := \widetilde{\xi}^Z_{\beta\beta}.$$

For a symmetric matrix $C = U\,\mathrm{Diag}(c)U^\top$, define its spectral positive and negative parts by $C_\pm := U\,\mathrm{Diag}(c_\pm)U^\top$. We construct certificate branches and compute their projections, obtaining both distance bounds and explicit normals for slack verification.

Certificate sets. To make limiting-normal certificates easier to compute, we construct a finite family of branches by imposing additional conditions on the zero spectral block of $\mathcal{K}_W(u)$. We begin with the two pure branches

$$\mathcal{B}_Y(u) := \{(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : \widetilde{\Gamma}^Y_{\beta\beta} = 0\},$$
$$\mathcal{B}_Z(u) := \{(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : \widetilde{\Gamma}^Z_{\beta\beta} = 0\},$$

and let

$$\mathcal{N}_{\mathrm{base}}(u) := \mathcal{K}_S(u) \cup \mathcal{B}_Y(u) \cup \mathcal{B}_Z(u).$$

For $\beta \neq \varnothing$, let $\mathcal{P}(\beta)$ be the set of all ordered partitions $\pi = (I_+, I_0, I_-)$ of $\{1, \dots, |\beta|\}$, with empty parts allowed. Choose finite sets $\mathcal{Q}(u;\xi) \subseteq \mathbb{O}^{|\beta|}$ and $\Theta_{\mathrm{cert}} \subseteq [0,1]$. The selected bases may depend on both u and ξ. For $\pi \in \mathcal{P}(\beta)$, $Q \in \mathcal{Q}(u;\xi)$, and $\theta \in \Theta_{\mathrm{cert}}$, write

$$G_Y := Q^\top \widetilde{\Gamma}^Y_{\beta\beta} Q, \qquad G_Z := Q^\top \widetilde{\Gamma}^Z_{\beta\beta} Q,$$

and define

$$\begin{aligned}\mathcal{B}_{\pi,Q,\theta}(u) := \big\{(\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : {} & (G_Y)_{++} = (G_Y)_{+0} = 0,\\ & (G_Y)_{00} \preceq 0, \quad (G_Z)_{00} \succeq 0,\\ & (G_Z)_{0-} = (G_Z)_{--} = 0,\\ & \theta(G_Y)_{+-} + (1-\theta)(G_Z)_{+-} = 0\big\}.\end{aligned}$$

Here the block subscripts refer to the partition π, and the transposed block conditions follow from symmetry. Set

$$\mathcal{N}_{\mathrm{cert}}(u;\xi) := \mathcal{N}_{\mathrm{base}}(u) \cup \bigcup_{\substack{\pi \in \mathcal{P}(\beta),\ Q \in \mathcal{Q}(u;\xi)\\ \theta \in \Theta_{\mathrm{cert}}}} \mathcal{B}_{\pi,Q,\theta}(u).$$

If $\beta = \varnothing$ or no mixed branches are selected, set $\mathcal{N}_{\mathrm{cert}}(u;\xi) := \mathcal{N}_{\mathrm{base}}(u)$.

Proposition 8 (Limiting-normal certificates) *For every $u \in \Omega$ and $\xi \in \mathcal{U}$,*

$$\mathcal{N}_{\mathrm{cert}}(u;\xi) \subseteq N_\Omega^{\lim}(u) = \mathcal{K}_M(u) \subseteq \mathcal{K}_W(u). \tag{76}$$

Consequently,

$$\mathrm{dist}(\xi, \mathcal{K}_W(u)) \le \mathrm{dist}(\xi, N_\Omega^{\lim}(u)) \le \mathrm{dist}\big(\xi, \mathcal{N}_{\mathrm{cert}}(u;\xi)\big). \tag{77}$$

If $\beta = \varnothing$, all sets in (76) *coincide, and both inequalities in* (77) *are equalities.*

Proof We use the spectral characterization in [9, Proposition 3.3 and Theorem 3.1], recalled in Appendix A. Fix $\pi = (I_+, I_0, I_-)$ and Q. Absorb a permutation of the partition into Q so that the positive, zero, and negative coordinates occur in this order. For $s \downarrow 0$, assign a common value $a_+(s)$ to the coordinates in I_+, zero to those in I_0, and $a_-(s)$ to those in I_-. Choose

	$a_+(s)$	$a_-(s)$	$a_+(s)/(a_+(s)-a_-(s))$
$0<\theta<1$	$s\theta$	$-s(1-\theta)$	θ
$\theta=0$	s^2	$-s$	$s/(1+s)$
$\theta=1$	s	$-s^2$	$1/(1+s)$

The resulting vectors $a(s)$ are ordered and converge to zero. Their divided-difference matrices converge to an element of $\mathbb{U}_{|\beta|}$ whose $+-$ block is the constant θ. The other blocks have exactly the values specified in Appendix A. Hence the degenerate-block conditions defining $\mathcal{B}_{\pi,Q,\theta}(u)$ satisfy that characterization, including the two endpoint values of θ.

The all-positive, all-negative, and all-zero partitions give $\mathcal{B}_Y(u)$, $\mathcal{B}_Z(u)$, and $\mathcal{K}_S(u)$, respectively. Empty parts impose no conditions. This proves the first inclusion in (76); the equality and the last inclusion follow from the same characterization. Distances reverse set inclusion, giving (77). When $\beta = \varnothing$, no degenerate-block restrictions remain, so all the sets coincide. □

For fixed (u, ξ) and selected bases, each branch is a nonempty closed convex cone. Comparing the branch projections therefore yields a closest normal in $\mathcal{N}_{\rm cert}(u;\xi)$. The number of branches is at most $3+|\mathcal{Q}(u;\xi)|\,|\Theta_{\rm cert}|\,3^{|\beta|}$.

Explicit distance formulas. To evaluate the resulting certificate distances, we first compute the contribution from the common linear block constraints. Write $d_W := \operatorname{dist}(\xi, \mathcal{K}_W(u))$. These constraints give

$$\begin{aligned} d_W^2 &= \|\widetilde{\xi}^Y_{\alpha\alpha}\|_F^2 + 2\|\widetilde{\xi}^Y_{\alpha\beta}\|_F^2 + \|\widetilde{\xi}^Z_{\gamma\gamma}\|_F^2 + 2\|\widetilde{\xi}^Z_{\beta\gamma}\|_F^2 \\ &\quad + 2\sum_{\substack{i\in\alpha\\ j\in\gamma}} \frac{\big(\Sigma_{ij}\widetilde{\xi}^Y_{ij} + (1-\Sigma_{ij})\widetilde{\xi}^Z_{ij}\big)^2}{\Sigma_{ij}^2 + (1-\Sigma_{ij})^2}. \end{aligned} \tag{78}$$

The basic certificate distance is

$$\begin{aligned} &\operatorname{dist}^2\big(\xi, \mathcal{N}_{\rm base}(u)\big) \\ &\quad = d_W^2 + \min\big\{\|(A_\beta)_+\|_F^2 + \|(B_\beta)_-\|_F^2,\ \|A_\beta\|_F^2,\ \|B_\beta\|_F^2\big\}. \end{aligned} \tag{79}$$

For a mixed branch, set $\widehat{A} := Q^\top A_\beta Q$ and $\widehat{B} := Q^\top B_\beta Q$. Then

$$\begin{aligned} &\operatorname{dist}^2\big(\xi, \mathcal{B}_{\pi,Q,\theta}(u)\big) \\ &\quad = d_W^2 + \|\widehat{A}_{++}\|_F^2 + 2\|\widehat{A}_{+0}\|_F^2 + \|(\widehat{A}_{00})_+\|_F^2 \\ &\qquad + \|\widehat{B}_{--}\|_F^2 + 2\|\widehat{B}_{0-}\|_F^2 + \|(\widehat{B}_{00})_-\|_F^2 \\ &\qquad + \frac{2\|\theta\widehat{A}_{+-} + (1-\theta)\widehat{B}_{+-}\|_F^2}{\theta^2 + (1-\theta)^2}. \end{aligned} \tag{80}$$

The squared distance to $\mathcal{N}_{\rm cert}(u;\xi)$ is the minimum of the squared distances to its included branches.

These formulas follow from blockwise orthogonal projections, with the factors 2 accounting for transposed off-diagonal blocks in the Frobenius norm. The same projections also provide an explicit normal witness for the slack verification below. For a pair of scalar entries or equally sized blocks (U_0, V_0) subject to $aU + bV = 0$, with $a^2 + b^2 > 0$, set

$$E := aU_0 + bV_0, \qquad U_* := U_0 - \frac{aE}{a^2+b^2}, \qquad V_* := V_0 - \frac{bE}{a^2+b^2}. \tag{81}$$

For the common $\alpha\gamma$ block, apply this formula entrywise with $(a,b) = (\Sigma_{ij}, 1-\Sigma_{ij})$. For a mixed $+-$ block, apply it with $(a,b) = (\theta, 1-\theta)$ to $(\widehat{A}_{+-}, \widehat{B}_{+-})$, and fill the transpose blocks by symmetry. Set all blocks required to vanish equal to zero and retain all free blocks.

For the basic branches, the projected $\beta\beta$ pair is $(-(A_\beta)_-, (B_\beta)_+)$, $(0, B_\beta)$, or $(A_\beta, 0)$ for $\mathcal{K}_S$, $\mathcal{B}_Y$, or $\mathcal{B}_Z$, respectively. For a mixed branch, use

$$(G_Y)_{00} = -(\widehat{A}_{00})_-, \qquad (G_Z)_{00} = (\widehat{B}_{00})_+.$$

Rotate the mixed degenerate block back by Q, then rotate the full pair back by P. These orthogonal transformations preserve the product Frobenius norm. Return the minimizing normal $\Gamma = (\Gamma^Y, \Gamma^Z)$, its branch type, the spectral data P and (α, β, γ), and, where applicable, Q, π, θ, together with $r = \|\xi - \Gamma\|$. The returned branch data permit an independent check of membership, while (79) or (80) gives the squared residual.

Slack verification. To use these certificates in the slack update, consider a feasible trial point $u^{t+1} \in \Omega$ and set

$$\xi^{t+1} := -\nabla_u \Phi^t(u^{t+1}) = (\rho + b_t)(\widehat{c}^t - u^{t+1}).$$

The computable test is

$$\operatorname{dist}\big(\xi^{t+1}, \mathcal{N}_{\mathrm{cert}}(u^{t+1}; \xi^{t+1})\big) \le \delta_u^t. \tag{82}$$

Let Γ^{t+1} be a closest certificate normal returned by the preceding block projections, attaining the distance in this test. When the test passes, Proposition 8 gives

$$\Gamma^{t+1} \in N_\Omega^{\lim}(u^{t+1}), \qquad \|\nabla_u \Phi^t(u^{t+1}) + \Gamma^{t+1}\| \le \delta_u^t.$$

Thus the slack residual condition (40) holds, and the candidate is accepted if it also satisfies the descent condition (39). Retain this normal and its branch data for evaluating the full inner residual in Lemma 5. At inner termination, the retained normal may also be used in the outer slack test whenever $\|\nabla_u L_k(x^{k+1}) + \Gamma^{k+1}\| \le \varepsilon_{\mathrm{stat}}$.

The candidate-generation routines and certificate choices used in the numerical experiments will be specified in Section 6.

6 Numerical experiments

We evaluate SPALM on two test families, denoted by Families I and II. We use two relative noise levels, $\eta_{\mathrm{rel}} \in \{0.01, 0.05\}$, and five seeds for each parameter group. Both families fit the SDCMPCC model (1) and its slack reformulation (2). We describe the implementation, evaluation criteria, models, and data generation below.

6.1 Implementation and evaluation criteria

We evaluate finite-budget runs using the final complete outer tuple, its updated multipliers, and normal witnesses constructed from retained joint spectral data. The formulas in Section 5.5 define the certificate calculations; Remark 6 specifies their interpretation in floating-point arithmetic. All matrix blocks use the Frobenius metric and all other blocks the Euclidean metric, with the canonical product norm. An isometric vectorization of symmetric matrices weights off-diagonal upper-triangular entries by $\sqrt{2}$.

Reported residuals. At a complete tuple $(w, Y, Z, \lambda, \mu, M^G, M^H)$, define

$$\begin{aligned}
r_{\mathrm{pri}} :=& \max\{\|g(w)_+\|_2, \|h(w)\|, \|G(w) - Y\|_F, \|H(w) - Z\|_F, \\
&\operatorname{dist}(Y, \mathbb{S}_+^n), \operatorname{dist}(Z, \mathbb{S}_-^n), \|YZ\|_F\}, \\
r_\lambda :=& \frac{\|(-\lambda)_+\|_2}{1 + \|\lambda\|_2}, \qquad r_{\lambda g} := \frac{\|\lambda \circ g(w)\|_2}{1 + \|\lambda\|_2 + \|g(w)\|_2}, \\
s_w :=& \nabla f(w) + g'(w)^*\lambda + h'(w)^*\mu + G'(w)^* M^G + H'(w)^* M^H, \\
r_{\mathrm{stat}} :=& \frac{\|s_w\|}{1 + \|\nabla f(w)\| + \|\lambda\|_2 + \|\mu\| + \|M^G\|_F + \|M^H\|_F}.
\end{aligned}$$

The inequality residuals are zero for Family II. The primal residual vanishes exactly at feasible points, while V_k in (22) instead controls the penalty update. Model-specific expressions for s_w are given below.

To evaluate the slack-stationarity component, we use the retained joint spectral representation of $u = (Y, Z)$. Following Section 5.5, the signs and structural zeros of the retained coefficients determine the spectral partition used for normal-witness evaluation. With $\xi = (M^G, M^H)$, let A_β and B_β denote its zero-block spectral coordinates in this representation.

The selected certificate family contains the three basic branches. For $1 \le |\beta| \le 6$, it also includes mixed branches constructed using

$$\mathcal{Q} = \{I, Q(A_\beta), Q(B_\beta), Q(A_\beta + B_\beta), Q(A_\beta - B_\beta)\}, \qquad \Theta_{\mathrm{cert}} = \{0, 1\},$$

where $Q(C)$ is a selected orthogonal eigenbasis and duplicate bases are discarded. All ordered partitions of the zero block are tested. For $|\beta| = 0$ or $|\beta| > 6$, the three basic branches are used.

At the same slack point $u = (Y, Z)$, the reporting witness Γ_{rep} minimizes $\|\xi - \Gamma\|$ over the retained inner witness and the projections onto the selected certificate branches. Define the normalization factor and the reporting residual by

$$s_N := 1 + \sqrt{\|M^G\|_F^2 + \|M^H\|_F^2}, \qquad r_{M,\Gamma} := \frac{\|\xi - \Gamma_{\mathrm{rep}}\|}{s_N},$$

and

$$R_{M,\Gamma} := \max\{r_{\mathrm{pri}}, r_\lambda, r_{\lambda g}, r_{\mathrm{stat}}, r_{M,\Gamma}\}.$$

A run is counted as successful when the absolute feasibility and stationarity tests associated with Algorithm 1, the reporting test $R_{M,\Gamma} \le 10^{-4}$, and the independent numerical checks are satisfied. The absolute feasibility and stationarity tolerances are both 10^{-4}. The absolute slack test in Algorithm 1 uses the retained inner normal, whereas Γ_{rep} is used for the reporting residual.

Tables report the final complete outer tuple and its updated multipliers. Here N is the number of planned runs and M is the number of successful runs. Each numerical column reports the median of its per-run values; $R_{M,\Gamma}$ is evaluated for each tuple before aggregation. Solver wall-clock time includes candidate generation, witness evaluation, and checkpoint storage. Independent-verification time is recorded separately.

Inner schedule and finite budgets. At outer iteration $k = 0, 1, \ldots$, we choose the inner initial point $x^{k,0}$ according to (23), using the fixed feasible reference specified below for each family. We set

$$a_t = b_t = 10^{-2}, \qquad \nu_{k+1} = \max\{10^{-6}, 10^{-1}(0.8)^{k+1}\}, \qquad \delta_w^t = \delta_u^t = \nu_{k+1}/4.$$

Each accepted block update is checked against the proximal descent condition and the corresponding absolute error test. The retained slack normal is used in the full residual evaluation. For the separated linear maps $G(w) = X$ and $H(w) = -S$, the inner return criterion is

$$\widehat{\mathcal{R}}_{\rho_k}^{t+1} = \left[\left(\rho_k\|\Delta u^t\| + a_t\|\Delta w^t\| + \delta_w^t\right)^2 + \left(b_t\|\Delta u^t\| + \delta_u^t\right)^2\right]^{1/2} \le \nu_{k+1}.$$

The outer and inner iteration limits are 400 and 200, respectively. We use $\rho_0 = 0.2$, penalty growth factor 1.5, and feasibility reduction factor 0.8. A requested penalty increase exceeding 30 terminates the run. Iteration, search, and time budgets have separately recorded termination statuses.

L-BFGS and conjugate gradients use iteration limits of 160 and 500, respectively, with a CG relative tolerance of 10^{-13}. The Huber–Stiefel search uses smoothing parameter 10^{-3}, gradient tolerance 10^{-10}, and an iteration limit of 480. Fixed-support polishing uses at most two initial bases, an iteration limit of 240, and gradient tolerance 10^{-10}. Three seeds are executed concurrently with one BLAS thread per process.

Slack-candidate generation. Slack candidates are generated by fixed-basis projections, Huber–Stiefel searches, and fixed-support polishing. For a proximal center (A, B), let $P = [P_+, P_-] \in \mathbb{O}^n$ have a fixed column partition. Compressed spectral projections give

$$Y = P_+ \Pi_{\mathbb{S}_+}(P_+^\top A P_+) P_+^\top, \qquad Z = P_- \Pi_{\mathbb{S}_-}(P_-^\top B P_-) P_-^\top,$$

where the projections act on symmetric matrices of the appropriate dimensions. The compressed eigendecompositions produce the joint spectral basis retained for normal-witness evaluation.

To improve the candidate while keeping the column partition fixed, fixed-support polishing varies P to maximize

$$F_{\mathrm{sup}}(P) = \|\Pi_{\mathbb{S}_+}(P_+^\top A P_+)\|_F^2 + \|\Pi_{\mathbb{S}_-}(P_-^\top B P_-)\|_F^2.$$

Each generated candidate is evaluated by the proximal descent and absolute normal-residual tests.

Remark 6 (Finite-precision implementation) The implementation uses finite search budgets, an accuracy floor $\nu_{\min} = 10^{-6}$, and an absolute descent-comparison tolerance of 10^{-11}. Structural spectral representations, normal witnesses, and residuals are checked in floating-point arithmetic. The reported verification is at the numerical-witness level; rigorous roundoff certification requires validated error enclosures. The selected certificate family provides a sufficient test for normal-residual accuracy. The asymptotic results in Sections 3–4 concern infinite sequences satisfying the stated vanishing-error assumptions.

6.2 Family I: a compact low-rank correlation formulation

Numerical methods for nearest low-rank correlation matrix problems include [23,41,6]. Given an observed symmetric matrix $C_{\text{obs}} \in \mathbb{S}^n$ and an integer $1 \le r \le n$, we consider

$$\begin{aligned} \min_{X \in \mathbb{S}^n} \quad & \frac{1}{2}\|X - C_{\text{obs}}\|_F^2 \\ \text{s.t.} \quad & \operatorname{diag}(X) = \mathbf{1}, \qquad X \succeq 0, \qquad \operatorname{rank}(X) \le r. \end{aligned} \tag{83}$$

The representation in [9] uses a trace equality and an additional semidefinite constraint on an auxiliary matrix. For a fixed lifting parameter $\vartheta \in (0,1)$, we use the following compact two-matrix SDCMPCC formulation:

$$\begin{aligned} \min_{X,S \in \mathbb{S}^n} \quad & \frac{1}{2}\|X - C_{\text{obs}}\|_F^2 \\ \text{s.t.} \quad & \operatorname{diag}(X) = \mathbf{1}, \\ & \|I - S\|_F^2 \le r + \vartheta, \\ & X \succeq 0, \qquad S \succeq 0, \qquad XS = 0. \end{aligned} \tag{84}$$

Set

$$w = (X,S), \quad g(w) = \|I - S\|_F^2 - r - \vartheta, \quad h(w) = \operatorname{diag}(X) - \mathbf{1}, \quad G(w) = X,\ H(w) = -S. \tag{85}$$

For this model, the reporting stationarity vector $s_w = (s_X, s_S)$ has the explicit blocks

$$\begin{aligned} s_X &= X - C_{\text{obs}} + \operatorname{Diag}(\mu) + M^G, \\ s_S &= 2\lambda(S - I) - M^H, \qquad \|s_w\|^2 = \|s_X\|_F^2 + \|s_S\|_F^2. \end{aligned}$$

The projection of the feasible set of (84) onto the X-space is exactly the feasible set of the rank-constrained correlation problem (83). Indeed, if $k = \operatorname{rank}(X)$, then $XS = 0$ implies $\operatorname{range}(X) \subseteq \ker S$. Hence S has at least k zero eigenvalues and

$$k \le \|I - S\|_F^2 \le r + \vartheta < r + 1.$$

Since k is an integer, this gives $k \le r$. Conversely, for any $X \succeq 0$ with $\operatorname{rank}(X) = k \le r$, let P_X be the orthogonal projector onto $\operatorname{range}(X)$ and set $S = I - P_X$. Then $S \succeq 0$, $XS = 0$, and $\|I - S\|_F^2 = k \le r < r + \vartheta$.

The lifted feasible set is also compact. Indeed, every feasible pair satisfies

$$\|X\|_F \le \operatorname{tr}(X) = n, \qquad \|S\|_F \le \sqrt{n} + \sqrt{r + \vartheta}.$$

The constraints are closed, and the set is nonempty, as shown by $X = \mathbf{1}\mathbf{1}^\top$ and $S = I - X/n$. The continuous objective therefore attains its minimum.

Remark 7 (Scope of the exact lifting) The preceding argument proves equality of the projected feasible sets. Since the objective is independent of S, the two formulations have the same global optimal value and corresponding global minimizers under projection. Lifted local minimizers need not project to local minimizers, and lifted M-stationarity need not imply stationarity with respect to feasible tangent directions of the original rank-constrained problem, because the auxiliary matrix S changes the local feasible geometry. Accordingly, M-stationarity and its residual estimates below concern the lifted SDCMPCC (84) (equivalently, its slack reformulation by Proposition 1), not an intrinsic stationarity notion for (83). A concrete counterexample is given in Appendix D.

Constraint qualification and inner level boundedness. We next verify the constraint qualification used in the outer analysis and the level-boundedness assumption for the inner subproblems.

Proposition 9 (Active-set SDCMPCC-LICQ for Family I) *The slack reformulation induced by* (84) *satisfies the active-set SDCMPCC-LICQ of Definition 5 at every feasible point. Equivalently, the unique slack lift* $((X,S),(X,-S))$ *of every feasible* (X,S) *satisfies this condition. This conclusion does not require strict complementarity or* $\operatorname{rank}(X) = r$.

The proof is given in Appendix C.1. The threshold $r + \vartheta \in (r, r+1)$ enforces the integer rank bound. At a feasible pair with S an orthogonal projector, $\|I - S\|_F^2 = n - \operatorname{rank}(S)$ is an integer, so the scalar inequality is strict.

Corollary 4 (Absence of singular M-multipliers for Family I) *For the constraint representation* (85) *of Family* I*, no feasible point of* (84) *admits a singular M-multiplier.*

The proof is given in Appendix C.1.

Remark 8 (Dependence on the constraint representation) The formulation in [9] uses an auxiliary matrix U satisfying

$$\operatorname{tr}(U) = r, \qquad U \succeq 0, \qquad (X, U - I) \in \Omega,$$

in addition to $\operatorname{diag}(X) = \mathbf{1}$. If $\operatorname{rank}(X) = r$, complementarity implies that U acts as the identity on $\operatorname{range}(X)$. Positive semidefiniteness and the trace equality then force $U = P_X$, where P_X is the orthogonal projector onto $\operatorname{range}(X)$. At this point, the pair $(X, U - I)$ is strictly complementary. Set the trace multiplier to $\tau = 1$, the diagonal-equality multiplier to zero, and

$$W = -(I - P_X), \qquad (\Gamma^G, \Gamma^H) = (0, -P_X).$$

Then $W \in N_{\mathbb{S}^n_+}(U)$ and

$$(\Gamma^G, \Gamma^H) \in \widehat{N}_\Omega(X, U - I) \subseteq N^{\lim}_\Omega(X, U - I),$$

as follows from the normal-cone formula in [9, Proposition 3.2]. Moreover,

$$\Gamma^G = 0, \qquad \tau I + W + \Gamma^H = 0.$$

Thus the earlier constraint representation admits a nonzero abnormal multiplier at every feasible point with $\operatorname{rank}(X) = r$. In contrast, Corollary 4 excludes singular M-multipliers at every feasible point of (84).

Proposition 10 (Inner level boundedness for Family I) *Fix $\rho > 0$ and any safeguarded multiplier vector Υ. For every x^0 in the inner domain, the sublevel set $\mathcal{S}_\rho(x^0)$ in (42) is bounded.*

The proof is given in Appendix C.1.

A feasible reference point for Family I. For $e = \mathbf{1} \in \mathbb{R}^n$, set

$$X^{\mathrm{ref}} = ee^\top, \qquad S^{\mathrm{ref}} = I - \frac{ee^\top}{n}, \qquad x^{\mathrm{ref}} = ((X^{\mathrm{ref}}, S^{\mathrm{ref}}), (X^{\mathrm{ref}}, -S^{\mathrm{ref}})). \tag{86}$$

This point is feasible: $\operatorname{diag} X^{\mathrm{ref}} = e$, $X^{\mathrm{ref}} S^{\mathrm{ref}} = 0$, both matrices are positive semidefinite, and $\|I - S^{\mathrm{ref}}\|_F^2 = 1 < r + \vartheta$ for $r \geq 1$. Lemma 2 therefore applies with $f(w^{\mathrm{ref}}) = \frac{1}{2}\|ee^\top - C_{\mathrm{obs}}\|_F^2$. For the inequality term, $g(w^{\mathrm{ref}}) = 1 - r - \vartheta < 0$ and $\bar{\lambda}^k \geq 0$ give $(g(w^{\mathrm{ref}}) + \bar{\lambda}^k/\rho_k)_+ \leq \bar{\lambda}^k/\rho_k$. The reference point is determined by the dimension n.

Planted instances and exact relative noise. We first construct a planted feasible pair $(X^\star, S^\star)$ with prescribed ranks and common kernel dimension, then perturb $X^\star$ to obtain the observations. Let $n \geq 2$ and choose integers r, d_0 satisfying $1 \leq r \leq n$ and $0 \leq d_0 \leq r - 1$. Set $r_0 := r - d_0$, so $1 \leq r_0 \leq r$. Generate $\widetilde{F} \in \mathbb{R}^{n \times r_0}$ with independent standard normal entries and normalize every row,

$$F_{i:} := \frac{\widetilde{F}_{i:}}{\left\|\widetilde{F}_{i:}\right\|_2}, \qquad X^\star := FF^\top.$$

With probability one, the normalization is well defined and preserves the full column rank of $\widetilde{F}$. Hence $X^\star \succeq 0$, $\operatorname{diag}(X^\star) = \mathbf{1}$, and $\operatorname{rank}(X^\star) = r_0$. Let $Q_X \in \mathbb{R}^{n \times r_0}$ have orthonormal columns spanning $\operatorname{range}(X^\star)$, and choose $Q_0 \in \mathbb{R}^{n \times d_0}$ with orthonormal columns in $\ker X^\star$. Define

$$S^\star := I - Q_X Q_X^\top - Q_0 Q_0^\top.$$

Consequently,

$$\operatorname{rank}(S^\star) = n - r, \qquad X^\star S^\star = 0, \qquad \|I - S^\star\|_F^2 = r,$$

and $\dim(\ker X^\star \cap \ker S^\star) = d_0$. The scalar inequality is strictly satisfied, with $g(X^\star, S^\star) = -\vartheta$. The parameter d_0 jointly specifies the planted common kernel dimension and rank $r_0 = r - d_0$. The spectral partition at termination is determined from the retained joint spectral representation.

To obtain observations at a prescribed relative noise level, draw $\widetilde{E} \in \mathbb{R}^{n\times n}$ with independent standard normal entries and set

$$\begin{aligned} E_0 &:= \frac{\widetilde{E}+\widetilde{E}^\top}{2} - \mathrm{Diag}\left(\mathrm{diag}\left(\frac{\widetilde{E}+\widetilde{E}^\top}{2}\right)\right), \\ E &:= \frac{E_0}{\|E_0\|_F}, \\ C_{\mathrm{obs}} &:= X^\star + \eta_{\mathrm{rel}}\,\|X^\star\|_F\,E, \qquad \eta_{\mathrm{rel}} \in \{0.01, 0.05\}. \end{aligned} \tag{87}$$

Since $n \geq 2$, E_0 is nonzero with probability one, so its normalization is well defined. It follows that

$$\mathrm{diag}(C_{\mathrm{obs}}) = \mathbf{1}, \qquad \frac{\|C_{\mathrm{obs}} - X^\star\|_F}{\|X^\star\|_F} = \eta_{\mathrm{rel}}.$$

For fixed $(n, r, d_0, \mathrm{seed})$, the same normalized direction E is used at both relative noise levels.

All Family I experiments use $\vartheta = 1/2$. The test grid is

$$(n, r) \in \{(20,5), (30,8), (50,10), (100,20)\}, \quad d_0 \in \{0, 2\}, \quad \eta_{\mathrm{rel}} \in \{0.01, 0.05\},$$

with five seeds for every parameter group. The experiments use the implementation settings in Section 6.1. The w-block is treated by L-BFGS with an analytic blockwise fallback. The per-run time limits are 5400 seconds. Time limits are checked at algorithmic checkpoints.

Table 1: Finite-budget results for Family I. Residuals and solver times are medians over five runs per parameter group; M/N denotes successful runs over planned runs. Time is the solver wall-clock time, including candidate generation, witness evaluation, and checkpoint storage.

d_0	n	r	η_{rel}	M/N	$R_{M,\Gamma}$	r_{stat}	$r_{M,\Gamma}$	r_{pri}	Time (s)
0	20	5	0.01	5/5	6.74×10^{-5}	6.74×10^{-5}	2.75×10^{-6}	2.84×10^{-5}	1.18
0	20	5	0.05	5/5	8.72×10^{-5}	1.27×10^{-5}	7.03×10^{-7}	8.72×10^{-5}	2.31
0	30	8	0.01	5/5	7.11×10^{-5}	7.11×10^{-5}	4.83×10^{-6}	4.21×10^{-5}	1.02
0	30	8	0.05	5/5	9.21×10^{-5}	2.39×10^{-5}	1.60×10^{-6}	9.21×10^{-5}	2.40
0	50	10	0.01	5/5	7.05×10^{-5}	7.05×10^{-5}	3.36×10^{-6}	4.76×10^{-5}	1.83
0	50	10	0.05	5/5	9.06×10^{-5}	5.92×10^{-6}	3.20×10^{-7}	9.06×10^{-5}	4.65
0	100	20	0.01	5/5	5.25×10^{-5}	5.25×10^{-5}	2.92×10^{-6}	3.21×10^{-5}	5.91
0	100	20	0.05	5/5	7.96×10^{-5}	4.22×10^{-6}	3.10×10^{-7}	7.96×10^{-5}	46.02
2	20	5	0.01	5/5	7.43×10^{-5}	6.77×10^{-5}	6.44×10^{-7}	7.43×10^{-5}	1.74
2	20	5	0.05	5/5	9.11×10^{-5}	7.85×10^{-6}	1.04×10^{-7}	9.11×10^{-5}	3.69
2	30	8	0.01	5/5	6.78×10^{-5}	6.43×10^{-5}	1.47×10^{-6}	6.08×10^{-5}	1.69
2	30	8	0.05	5/5	9.65×10^{-5}	8.69×10^{-6}	2.53×10^{-7}	9.65×10^{-5}	3.41
2	50	10	0.01	5/5	8.06×10^{-5}	5.23×10^{-5}	1.15×10^{-6}	8.06×10^{-5}	2.57
2	50	10	0.05	5/5	9.02×10^{-5}	5.64×10^{-6}	1.34×10^{-7}	9.02×10^{-5}	5.34
2	100	20	0.01	5/5	8.60×10^{-5}	3.24×10^{-5}	1.62×10^{-6}	8.60×10^{-5}	6.90
2	100	20	0.05	5/5	8.64×10^{-5}	4.10×10^{-6}	2.05×10^{-7}	8.64×10^{-5}	59.44

6.3 Family II: inverse linear SDP with a primal observation

Applications of inverse conic programming include portfolio selection and utility identification [18]. Inverse SDP models also arise in structural design: Ghatrani and Ghate [14] study the recovery of constraint matrices and include inverse truss design among their applications. We use an observation-fitting inverse SDP with fixed constraint data, as described below.

Given an observed cost matrix $\widehat{C} \in \mathbb{S}^n$ and a primal observation $X_{\mathrm{obs}} \in \mathbb{S}^n$, we fit a cost matrix and a corresponding primal optimizer of the forward SDP $\min\{\langle C, U\rangle : \langle A, U\rangle = 1,\ U \succeq 0\}$. Following the minimum-distance approach to noisy observations [5,8], the objective penalizes deviations from $\widehat{C}$ and X_{obs}. This fixes

the constraint data, unlike left-hand-side recovery in [14]; related inverse SDCMPCC formulations also adjust right-hand sides [37].

The dual of the forward SDP is

$$\max_{y\in\mathbb{R},\ S\in\mathbb{S}^n} \ y \quad \text{s.t.} \quad C - yA = S \succeq 0,$$

where y is the scalar equality multiplier and S is the dual slack. Write $A = \overline{A}/b_0$, with $b_0 > 0$ and $\overline{A}$ indefinite. The working dual coordinate $\widetilde{y} = y/b_0$ converts $yA + S = C$ into $C - \widetilde{y}\overline{A} - S = 0$, giving

$$\begin{aligned} \min_{C,X,S,\widetilde{y}} \quad & \frac{1}{2}\left\|C - \widehat{C}\right\|_F^2 + \frac{1}{2}\|X - X_{\text{obs}}\|_F^2 \\ \text{s.t.} \quad & \langle A, X\rangle = 1, \\ & C - \widetilde{y}\,\overline{A} - S = 0, \\ & X \succeq 0, \qquad S \succeq 0, \qquad XS = 0. \end{aligned} \tag{88}$$

To express (88) in the form of (1), set

$$w = (C, X, S, \widetilde{y}), \qquad h(w) = \big(\langle A, X\rangle - 1,\ C - \widetilde{y}\,\overline{A} - S\big).$$

Consistently with the sign convention in (1), define

$$G(w) = X, \qquad H(w) = -S. \tag{89}$$

Writing the equality multiplier as $\mu = (\sigma, T) \in \mathbb{R} \times \mathbb{S}^n$, the stationarity blocks are

$$\begin{aligned} s_C &= C - \widehat{C} + T, \quad s_X = X - X_{\text{obs}} + \sigma A + M^G, \\ s_S &= -T - M^H, \quad\ s_{\widetilde{y}} = -\langle \overline{A}, T\rangle. \end{aligned}$$

Thus

$$\begin{aligned} \|s_w\|^2 &= \|s_C\|_F^2 + \|s_X\|_F^2 + \|s_S\|_F^2 + |s_{\widetilde{y}}|^2, \\ \|h(w)\|^2 &= (\langle A, X\rangle - 1)^2 + \|C - \widetilde{y}\overline{A} - S\|_F^2, \qquad \|\mu\|^2 = \sigma^2 + \|T\|_F^2. \end{aligned}$$

In physical coordinates, the other stationarity blocks are unchanged and

$$s_y = -\langle A, T\rangle = \frac{s_{\widetilde{y}}}{b_0}, \qquad \|s_w^{\text{phys}}\|^2 = \|s_C\|_F^2 + \|s_X\|_F^2 + \|s_S\|_F^2 + b_0^{-2}|s_{\widetilde{y}}|^2.$$

For each fixed $b_0 > 0$, this invertible change of variables preserves exact stationarity. A common finite tolerance in working coordinates does not give a common physical-coordinate tolerance across instances with small b_0.

The forward single-equality SDP is strictly feasible. To see this, choose a unit eigenvector v of A with eigenvalue $\alpha > 0$ and take $\epsilon > 0$ small enough that $1 - \epsilon\,\mathrm{tr}(A) > 0$. Then

$$U_\epsilon := \epsilon I + \frac{1 - \epsilon\,\mathrm{tr}(A)}{\alpha} vv^\top \succ 0, \qquad \langle A, U_\epsilon\rangle = 1.$$

By Slater's theorem, a finite optimal value of the forward problem is attained by its dual. Every forward optimal solution can therefore be paired with y and $S = C - yA \succeq 0$ satisfying $XS = 0$. Thus the KKT representation is complete for this family whenever a forward optimum exists. Strict feasibility alone does not guarantee a finite forward optimal value for every cost matrix C.

Constraint qualification and inner level boundedness.

Proposition 11 (Active-set SDCMPCC-LICQ for Family II) *The unique slack lift* $((C, X, S, \widetilde{y}), (X, -S))$ *of every feasible point of* (88) *satisfies the active-set SDCMPCC-LICQ of Definition 5.*

The proof in Appendix C.2 uses the single equality $\langle A, X\rangle = 1$ and does not require A to be indefinite. The following level-boundedness result uses indefiniteness.

Proposition 12 (Inner level boundedness for Family II) *Suppose that* $\overline{A}$ *is indefinite. Fix* $\rho > 0$ *and any safeguarded multiplier vector* Υ*. For every* x^0 *in the inner domain, the sublevel set* $\mathcal{S}_\rho(x^0)$ *in* (42) *is bounded.*

The proof is given in Appendix C.2.

A feasible reference point for Family II. Choose a unit eigenvector v of A with $Av = \alpha v$ and $\alpha > 0$, and set

$$X^{\mathrm{ref}} = \frac{vv^\top}{\alpha}, \qquad C^{\mathrm{ref}} = S^{\mathrm{ref}} = 0, \qquad \widetilde{y}^{\mathrm{ref}} = 0. \tag{90}$$

Then $\langle A, X^{\mathrm{ref}}\rangle = 1$, the matrix equality holds, and the complementarity conditions hold. Pairing these variables with $(Y^{\mathrm{ref}}, Z^{\mathrm{ref}}) = (X^{\mathrm{ref}}, 0)$ gives a feasible reference for (2), with $f(w^{\mathrm{ref}}) = \frac{1}{2}\|\widehat{C}\|_F^2 + \frac{1}{2}\|vv^\top/\alpha - X_{\mathrm{obs}}\|_F^2$. A positive eigenvalue suffices for this construction. Indefiniteness is still used to bound the remaining variables in the model-level convergence result below.

The model bounds and the reference-point comparison give the following conclusion for both families.

Corollary 5 (Outer boundedness with feasible-reference comparison) *Fix an instance of Family* I *or Family* II*, with $\overline{A}$ indefinite in the latter case. Consider an infinite run of Algorithm 1 with Step 1 omitted, completed Step 3 solves, and $\nu_k \to 0$. Supply the corresponding feasible reference point in* (86) *or* (90) *and enforce* (23) *at every outer iteration. Then the primal outer sequence is bounded and has a nonempty set of accumulation points. Every accumulation point is M-stationary for the corresponding SDCMPCC and its slack reformulation. If Step 1 is retained with positive stopping tolerances, the algorithm terminates after finitely many outer iterations, provided every inner solve completes.*

The proof is given in Appendix C.3. A common AL upper bound yields the stated convergence conclusions, and Lemma 2 provides this bound under the feasible-reference comparison. For Family I, the stationarity conclusion concerns the lifting, with the scope described in Remark 7. In all reported runs, $L_k(x^k) < L_k(x^{\mathrm{ref}})$ at every outer iteration, so $x^{k,0} = x^k$. Accepted inner updates satisfy $L_k(x^{k+1}) \le L_k(x^{k,0})$.

Propositions 10 and 12 also ensure that each AL subproblem attains a global minimum: its initial sublevel set is nonempty and compact, and points outside it have larger objective values. At a minimizer $x_* = (w_*, u_*)$, the Fermat rule gives

$$\nabla_w L_k(x_*) = 0, \qquad -\nabla_u L_k(x_*) \in \widehat{N}_\Omega(u_*) \subseteq N_\Omega^{\lim}(u_*).$$

Thus a point satisfying Step 3 exists for every positive target. Proposition 5 and Corollary 2 further give finitely many accepted transitions under ideal block oracles and the stated parameter and error conditions. These existence statements do not guarantee that the practical routines return a candidate passing both acceptance tests within the allotted budget.

Planted instances and exact relative noise. For integers $n \ge 2$ and $1 \le r \le n-1$, draw an orthogonal matrix $Q = [Q_X, Q_S]$, with $Q_X \in \mathbb{R}^{n\times r}$ and $Q_S \in \mathbb{R}^{n\times(n-r)}$, and use positive eigenvalues geometrically spaced from 1 to 0.1 to construct complementary matrices

$$X^\star = Q_X D_X Q_X^\top, \qquad \overline{S} = Q_S D_S Q_S^\top,$$

which are subsequently normalized to unit Frobenius norm. Thus, $X^\star\overline{S} = 0$, $\mathrm{rank}(X^\star) = r$, and $\mathrm{rank}(\overline{S}) = n-r$. Both normalizations have nonzero denominators. Independently generate an indefinite matrix $\overline{A}$ with $\|\overline{A}\|_F = 1$. The construction requires $\langle \overline{A}, X^\star\rangle \neq 0$; resample if this inner product vanishes. Otherwise, change the sign of $\overline{A}$ if necessary so that $b_0 := \langle \overline{A}, X^\star\rangle > 0$, and set

$$A = \frac{\overline{A}}{b_0}, \qquad \kappa = \left\|\overline{A} + \overline{S}\right\|_F,$$

$$C^\star = \frac{\overline{A} + \overline{S}}{\kappa}, \qquad S^\star = \frac{\overline{S}}{\kappa}, \qquad \widetilde{y}^\star = \frac{1}{\kappa}, \qquad y^\star = b_0\widetilde{y}^\star = \frac{b_0}{\kappa}.$$

Here $\kappa > 0$: otherwise $\overline{A} = -\overline{S} \preceq 0$, contradicting indefiniteness. It follows that $\langle A, X^\star\rangle = 1$, $C^\star - \widetilde{y}^\star\overline{A} - S^\star = 0$, $X^\star S^\star = 0$, and $\|C^\star\|_F = \|X^\star\|_F = 1$. Equivalently, the physical dual multiplier satisfies $y^\star A + S^\star = C^\star$.

To construct the observed cost matrix, draw $\widetilde{E}_{ij} \overset{\mathrm{iid}}{\sim} \mathcal{N}(0,1)$ and set

$$E_0 = \frac{\widetilde{E} + \widetilde{E}^\top}{2} - \left\langle \frac{\widetilde{E} + \widetilde{E}^\top}{2}, \overline{A} \right\rangle \overline{A}, \qquad E = \frac{E_0}{\|E_0\|_F}.$$

Since $\|\overline{A}\|_F = 1$, this construction gives $\langle E_0, \overline{A}\rangle = 0$. For $n \ge 2$, $E_0 \neq 0$ with probability one, and hence $\|E\|_F = 1$ and $\langle E, \overline{A}\rangle = 0$. The observations are

$$\widehat{C} = C^\star + \eta_{\mathrm{rel}} E, \qquad X_{\mathrm{obs}} = X^\star, \qquad \eta_{\mathrm{rel}} \in \{0.01, 0.05\}. \tag{91}$$

Because $\|C^\star\|_F = \|E\|_F = 1$,

$$\frac{\left\|\widehat{C} - C^\star\right\|_F}{\|C^\star\|_F} = \eta_{\mathrm{rel}}.$$

The noise affects only the objective observation, so the planted primal–dual point remains feasible for every relative noise level.

We use $(n, r) \in \{(20, 5), (30, 8), (50, 10), (100, 20)\}$ and five seeds for each $(n, r, \eta_{\mathrm{rel}})$ group. The experiments use the implementation settings in Section 6.1 and a per-run time limit of 7200 seconds. The quadratic w-block is solved by matrix-free conjugate gradients with absolute block-residual tests.

Table 2: Finite-budget results for Family II. Each numerical column reports the median over five final complete outer tuples. M/N denotes successful runs over planned runs. Time is the solver wall-clock time, including candidate generation, witness evaluation, and checkpoint storage.

n	r	η_{rel}	M/N	$R_{M,\Gamma}$	r_{stat}	$r_{M,\Gamma}$	r_{pri}	Time (s)
20	5	0.01	5/5	8.67×10^{-5}	8.67×10^{-5}	8.64×10^{-6}	1.91×10^{-6}	4.09
20	5	0.05	5/5	7.44×10^{-5}	2.56×10^{-5}	6.76×10^{-5}	1.82×10^{-5}	2.34
30	8	0.01	5/5	8.71×10^{-5}	8.71×10^{-5}	6.75×10^{-7}	1.25×10^{-6}	9.65
30	8	0.05	5/5	8.57×10^{-5}	8.13×10^{-5}	3.00×10^{-7}	4.82×10^{-6}	5.76
50	10	0.01	5/5	8.61×10^{-5}	8.61×10^{-5}	3.02×10^{-5}	2.16×10^{-6}	42.03
50	10	0.05	5/5	8.09×10^{-5}	8.09×10^{-5}	2.95×10^{-5}	2.30×10^{-6}	40.20
100	20	0.01	5/5	8.38×10^{-5}	8.38×10^{-5}	3.02×10^{-5}	3.15×10^{-6}	82.42
100	20	0.05	5/5	8.12×10^{-5}	7.97×10^{-5}	2.95×10^{-5}	3.48×10^{-6}	91.41

For Family II, we also report

$$e_C := \frac{\|C - C^\star\|_F}{\|C^\star\|_F}, \qquad e_X := \frac{\|X - X^\star\|_F}{\|X^\star\|_F}.$$

Both denominators are one. The quantity e_C measures the relative deviation from the clean cost matrix, while e_X measures agreement with the observed primal target $X_{\mathrm{obs}} = X^\star$. The following rounded ranges summarize groupwise medians across the four tested dimensions:

η_{rel}	e_C	e_X
0.01	$[7.70, 8.26] \times 10^{-3}$	$[1.82, 2.15] \times 10^{-3}$
0.05	$[3.85, 4.13] \times 10^{-2}$	$[9.11 \times 10^{-3}, 1.08 \times 10^{-2}]$

6.4 Discussion of the results

All 120 planned runs satisfied the success criteria in Section 6.1 within the prescribed budgets: 80 for Family I and 40 for Family II (Tables 1 and 2). Across parameter groups, the median $R_{M,\Gamma}$ ranges from 5.25×10^{-5} to 9.65×10^{-5} for Family I and from 7.44×10^{-5} to 8.71×10^{-5} for Family II.

Figures 1 and 2 summarize the final residuals and solver times. Median solver times range from 1.02 to 59.44 seconds for Family I and from 2.34 to 91.41 seconds for Family II. For Family I, the median time is larger at $\eta_{\mathrm{rel}} = 0.05$ than at $\eta_{\mathrm{rel}} = 0.01$ for every tested (n, r, d_0). Family II exhibits no consistent ordering of solver time between the two noise levels.

For Family II, the group medians of both e_C and e_X are larger at $\eta_{\mathrm{rel}} = 0.05$ than at $\eta_{\mathrm{rel}} = 0.01$ in all four tested dimensions, as reported in Section 6.3. Here e_C measures deviation from the clean cost matrix, while e_X measures agreement with the observed primal target $X_{\mathrm{obs}} = X^\star$.

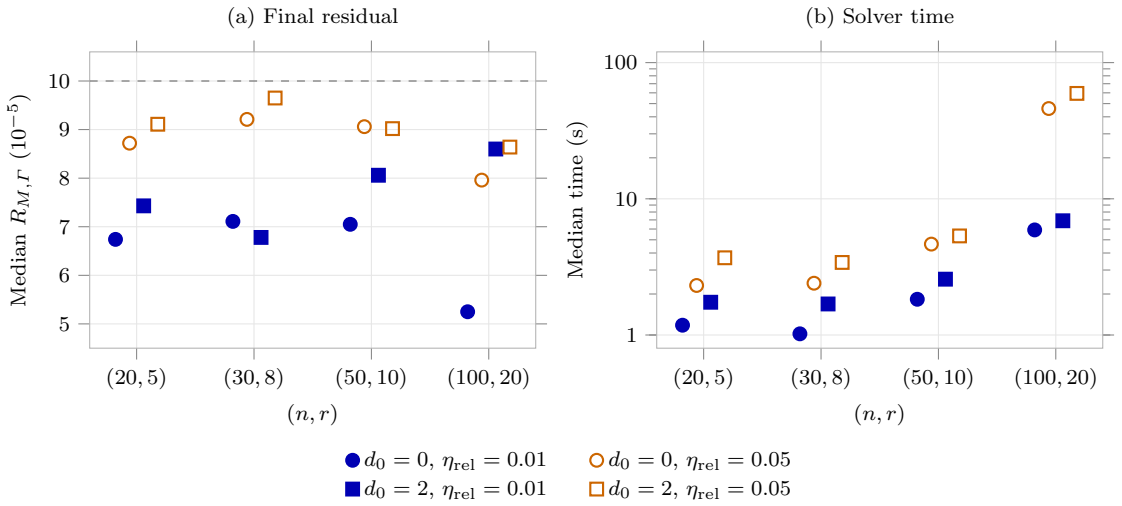


Fig. 1: Final residuals and solver times for Family I, using the group medians in Table 1. Each marker summarizes five runs at the indicated (n, r), noise level, and d_0. Residuals are displayed in units of 10^{-5}; the dashed line corresponds to $R_{M,\Gamma} = 10^{-4}$. The time axis is logarithmic.

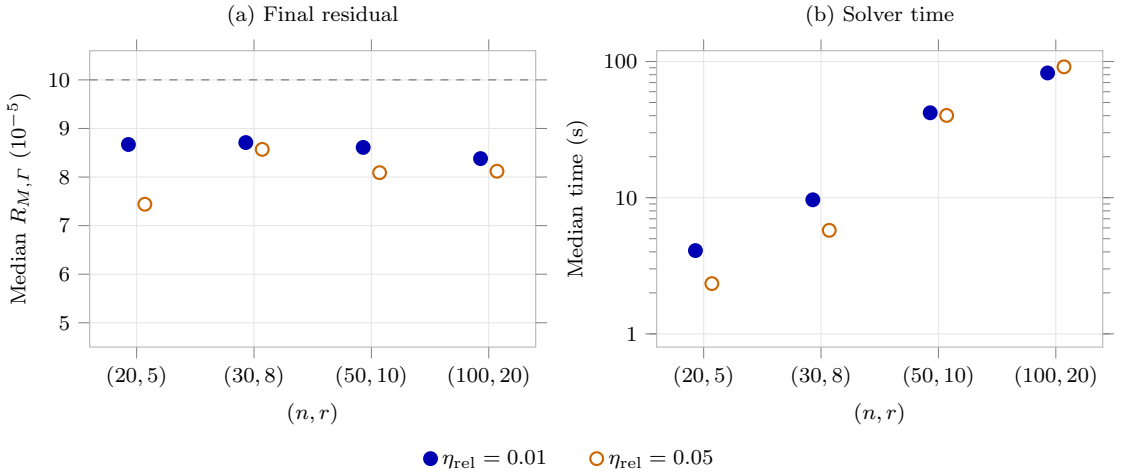


Fig. 2: Final residuals and solver times for Family II, using the group medians in Table 2. Each marker summarizes five runs at the indicated (n, r) and noise level. The residual scale, reporting threshold, and logarithmic time axis are as in Figure 1.

7 Conclusion

SPALM applies the safeguarded augmented Lagrangian framework of Jia et al. [20] to SDCMPCCs with a subproblem accuracy target defined by the limiting normal cone to Ω. This target includes coupled first-order normals that a scalar orthogonality representation can omit even at strictly complementary points. Feasible slack candidates enforce complementarity, while smooth linking residuals control the distance of the matrix image to Ω. For infinite outer sequences with vanishing absolute subproblem errors, feasible accumulation points are AM-stationary and, under AM-regularity, M-stationary. Active-set SDCMPCC-LICQ is sufficient for this regularity. When solving an augmented Lagrangian subproblem, the proximal alternating framework gives convergence of the full residual to zero along infinite accepted sequences under the stated level-boundedness, proximal-parameter, and error conditions. Square-summable block errors yield an $O(T^{-1/2})$ best-residual bound in accepted iterations.

Huber–Stiefel and fixed-basis procedures generate feasible slack candidates, whose acceptance requires descent tests and limiting-normal certificates. Spectral bounds quantify the objective gap of the fixed-basis candidates, but finite-budget candidate generation need not produce an accepted update. The correlation-matrix lifting has a compact feasible set and satisfies active-set SDCMPCC-LICQ at every feasible point. It preserves the projected feasible set and global optimal value. Its local geometry differs from that of the rank-constrained

problem: a lifted M-stationary local minimizer can project to a point with a feasible first-order descent direction, as shown in Appendix D.

For the two model families, explicit feasible references and the comparison rule (23) give a uniform outer AL upper bound. Under the stated model assumptions, completed inner solves, and vanishing absolute errors, the resulting outer sequence is bounded and all its accumulation points are M-stationary.

The finite-budget experiments report successful termination in all 120 runs for the lifting and a single-equality observation-fitting inverse SDP. Success combines absolute feasibility and stationarity tests with a normalized residual evaluated using a retained normal witness. These floating-point checks provide numerical evidence for the tested implementation. Further work should quantify the work required to generate accepted updates and develop validated error bounds for the normal-witness calculations.

A Complete spectral characterization of $\mathcal{K}_M$

This appendix gives the full spectral characterization of $\mathcal{K}_M(u) = N_\Omega^{\lim}(u)$ from [9, Proposition 3.3 and Theorem 3.1]. Fix $u = (Y, Z) \in \Omega$ and use the joint decomposition, spectral coordinates, and common multiplier space $\mathcal{K}_W(u)$ introduced in Subsection 2.1.

Let β be the zero spectral index set associated with $u = (Y, Z) \in \Omega$. Within the zero block, identify β with the local indices $\{1, \dots, |\beta|\}$. Let $\mathcal{P}(\beta)$ consist of ordered triples of pairwise disjoint subsets whose union is β, with empty subsets allowed. Write

$$\mathbb{R}^{|\beta|}_{\geq} := \left\{ a \in \mathbb{R}^{|\beta|} : a_1 \geq \cdots \geq a_{|\beta|} \right\}.$$

For $a \in \mathbb{R}^{|\beta|}_{\geq}$, let $D(a)$ denote the generalized first divided-difference matrix of $t \mapsto \max\{t, 0\}$ at a, namely,

$$(D(a))_{ij} = \begin{cases} \dfrac{\max\{a_i, 0\} - \max\{a_j, 0\}}{a_i - a_j}, & a_i \neq a_j, \\ 1, & a_i = a_j > 0, \\ 0, & a_i = a_j \leq 0, \end{cases} \qquad i, j = 1, \dots, |\beta|.$$

Define

$$\mathbb{U}_{|\beta|} := \left\{ \overline{\Xi} \in \mathbb{S}^{|\beta|} : \overline{\Xi} = \lim_{k \to \infty} D(a^k),\ a^k \to 0,\ a^k \in \mathbb{R}^{|\beta|}_{\geq} \right\}.$$

For any $\Xi_1 \in \mathbb{U}_{|\beta|}$, choose a sequence a^k realizing its divided-difference limit. After passing to a subsequence, the sign sets

$$\beta_+ := \{i : a_i^k > 0\}, \qquad \beta_0 := \{i : a_i^k = 0\}, \qquad \beta_- := \{i : a_i^k < 0\}$$

are independent of k. We call the resulting partition associated with Ξ_1; the same matrix Ξ_1 may admit more than one associated partition. Since the vectors a^k are ordered, these sets occur in positive, zero, and negative order. Thus there is a partition

$$\pi(\beta) := (\beta_+, \beta_0, \beta_-) \in \mathcal{P}(\beta)$$

such that

$$\Xi_1 = \begin{bmatrix} E_{\beta_+\beta_+} & E_{\beta_+\beta_0} & (\Xi_1)_{\beta_+\beta_-} \\ E_{\beta_+\beta_0}^\top & 0 & 0 \\ (\Xi_1)_{\beta_+\beta_-}^\top & 0 & 0 \end{bmatrix},$$

where every entry of $(\Xi_1)_{\beta_+\beta_-}$ belongs to $[0, 1]$.

For this choice of the realizing sequence and partition, define

$$\Xi_2 := \begin{bmatrix} 0 & 0 & E_{\beta_+\beta_-} - (\Xi_1)_{\beta_+\beta_-} \\ 0 & 0 & E_{\beta_0\beta_-} \\ \left(E_{\beta_+\beta_-} - (\Xi_1)_{\beta_+\beta_-}\right)^\top & E_{\beta_0\beta_-}^\top & E_{\beta_-\beta_-} \end{bmatrix}.$$

For $Q \in \mathbb{O}^{|\beta|}$, Q_{β_0} denotes the submatrix of Q formed by the columns corresponding to the index set β_0.

With these objects, the limiting normal cone has the following spectral representation:

$$\begin{aligned} \mathcal{K}_M(u) = \Big\{ (\Gamma^Y, \Gamma^Z) \in \mathcal{K}_W(u) : & \text{ there exist } Q \in \mathbb{O}^{|\beta|} \text{ and } \Xi_1 \in \mathbb{U}_{|\beta|} \text{ with an associated partition} \\ & \pi(\beta) = (\beta_+, \beta_0, \beta_-) \text{ and matrix } \Xi_2 \text{ such that} \\ & \Xi_1 \circ \left(Q^\top \widetilde{\Gamma}^Y_{\beta\beta} Q\right) + \Xi_2 \circ \left(Q^\top \widetilde{\Gamma}^Z_{\beta\beta} Q\right) = 0, \\ & Q_{\beta_0}^\top \widetilde{\Gamma}^Y_{\beta\beta} Q_{\beta_0} \preceq 0, \qquad Q_{\beta_0}^\top \widetilde{\Gamma}^Z_{\beta\beta} Q_{\beta_0} \succeq 0 \Big\}. \end{aligned}$$

All conditions involving empty index sets are vacuous. If $\beta = \varnothing$, the degenerate-block conditions disappear and $\mathcal{K}_M(u) = \mathcal{K}_W(u)$. If $Y + Z = 0$, then $Y = Z = 0$ and $\mathcal{K}_W(u) = \mathbb{S}^n \times \mathbb{S}^n$; the formula reduces to the limiting normal cone at the origin. The union in this characterization ranges over all such choices of Q, Ξ_1, and associated partition. Both the divided-difference limit and its associated sign partition are required; entrywise bounds on the mixed block alone do not characterize the admissible choices.

B Proof of the fixed-basis objective-gap bound

Proof (Proof of Proposition 7) Set $\mathcal{D} := \begin{pmatrix} A_+^2 & 0 \\ 0 & B_-^2 \end{pmatrix} \in \mathbb{S}^{2n}$, with A_+, B_- the spectral positive and negative parts. For each column p_i of $P \in \mathbb{O}^n$, choose $q_i = (p_i, 0)$ if $(p_i^\top A p_i)_+ \geq (p_i^\top B p_i)_-$, and $q_i = (0, p_i)$ otherwise. The vectors q_i are orthonormal. Jensen's inequality for $t \mapsto (t_+)^2$, applied in an eigenbasis of A or $-B$, gives

$$F(P) \leq \sum_{i=1}^{n} q_i^\top \mathcal{D} q_i \leq \sum_{j=1}^{n} \lambda_j^{\downarrow}(\mathcal{D}) = U_{\mathrm{spec}}.$$

For the second inequality, write the rank-n orthogonal projector $\sum_i q_i q_i^\top$ in an eigenbasis of $\mathcal{D}$: its diagonal entries lie in $[0,1]$ and sum to n, so their weighted sum is at most the sum of the n largest eigenvalues. Equation (60) now gives $L_{\mathrm{spec}} \leq v_\star \leq V_A$.

To obtain (74), select n entries attaining U_{spec} from the two spectral lists. If k come from the A-list, pair these with the k unselected B entries, and pair the other A entries with the selected B entries. Each selected entry is at least its partner, so the sum of maxima is U_{spec}. Ties are arbitrary but the orderings of P_B and $\lambda(B)$ must agree.

Finally, $\mathrm{diag}(P_A^\top B P_A) = T\lambda(B)$. On $[-R_0, R_0]$, the functions $\phi_i(s) := \max\{(\lambda_i(A))_+^2, (s_-)^2\}$ are $2R_0$-Lipschitz. Hence

$$\begin{aligned} V_A - L_{\mathrm{spec}} &= \sum_i \big[\phi_i(\lambda_i(B)) - \phi_i((T\lambda(B))_i)\big] \\ &\leq 2R_0 \|(I-T)\lambda(B)\|_1 \leq 2\sqrt{n}\, R_0 \|I - T\|_2 \|B\|_F. \end{aligned} \tag{92}$$

Together with $L_{\mathrm{spec}} \leq v_\star \leq V_A$, this proves the claim. □

C Verification of the assumptions for the numerical models

C.1 Family I

Proof (Proof of Proposition 9) Let $((\bar{X}, \bar{S}), (\bar{Y}, \bar{Z})) \in \widetilde{\mathcal{F}}$ be a feasible point of the induced slack reformulation. The linking equalities give $\bar{Y} = \bar{X}$ and $\bar{Z} = -\bar{S}$. By Proposition 2, consider a scalar λ, a vector $\mu \in \mathbb{R}^n$, and matrices $\Gamma^Y, \Gamma^Z \in \mathbb{S}^n$ satisfying the homogeneous reduced system. With $D := \mathrm{Diag}(\mu)$, its two stationarity equations and spectral relation are

$$D + \Gamma^Y = 0, \qquad 2\lambda(\bar{S} - I) - \Gamma^Z = 0, \qquad \bar{X}\Gamma^Y + \Gamma^Z \bar{S} = 0,$$

where $\lambda = 0$ whenever the scalar inequality is inactive and λ is unrestricted when it is active. Eliminating Γ^Y and Γ^Z gives

$$-\bar{X}D + 2\lambda(\bar{S} - I)\bar{S} = 0.$$

Since $\bar{X}\bar{S} = 0$, the two terms are orthogonal because

$$\langle \bar{X}D, (\bar{S} - I)\bar{S} \rangle = \mathrm{tr}\big(D\bar{X}(\bar{S} - I)\bar{S}\big) = 0.$$

Hence

$$\|\bar{X}D\|_F^2 + 4\lambda^2 \|(\bar{S} - I)\bar{S}\|_F^2 = 0.$$

Thus $\bar{X}D = 0$; since $\mathrm{diag}(\bar{X}) = \mathbf{1}$, $(\bar{X}D)_{ii} = \bar{X}_{ii}\mu_i = \mu_i$, and hence $\mu = 0$. If the inequality is inactive, then $\lambda = 0$. If it is active and $\lambda \neq 0$, then $(\bar{S} - I)\bar{S} = 0$, so $\bar{S}$ is an orthogonal projector. Consequently $\|I - \bar{S}\|_F^2$ is an integer, contradicting $\|I - \bar{S}\|_F^2 = r + \vartheta \in (r, r+1)$. Therefore $\lambda = 0$, and the stationarity equations yield $\Gamma^Y = \Gamma^Z = 0$. □

Proof (Proof of Corollary 4) Proposition 9 excludes nonzero homogeneous tuples in the full linearized multiplier space, and hence in the smaller set of actual normal-cone multipliers. □

Proof (Proof of Proposition 10) Fix $\rho > 0$ and the multipliers. Let $c := L_\rho(x^0)$ and take any $x = (X, S, Y, Z) \in \mathcal{S}_\rho(x^0)$. All terms in (20) are nonnegative. The objective therefore bounds X, while

$$\left(g(X,S) + \frac{\lambda}{\rho}\right)_+ \leq \sqrt{\frac{2c}{\rho}}$$

implies

$$\|I - S\|_F^2 \leq r + \vartheta + \sqrt{\frac{2c}{\rho}} - \frac{\lambda}{\rho}.$$

Thus S is bounded. Keeping the multiplier shifts in the linking terms gives

$$\|X - Y\|_F \leq \sqrt{\frac{2c}{\rho}} + \frac{\|M^G\|_F}{\rho}, \qquad \|S + Z\|_F \leq \sqrt{\frac{2c}{\rho}} + \frac{\|M^H\|_F}{\rho}.$$

Hence Y and Z are bounded as well. These bounds depend on ρ and the multipliers. □

C.2 Family II

Proof (Proof of Proposition 11) Let $(\bar{w}, (\bar{Y}, \bar{Z})) \in \widetilde{\mathcal{F}}$ be feasible for the induced slack reformulation, where $\bar{w} = (\bar{C}, \bar{X}, \bar{S}, \bar{\widetilde{y}}_0)$. Thus $\bar{Y} = \bar{X}$ and $\bar{Z} = -\bar{S}$. In the reduced system of Proposition 2, write the equality multiplier as $(\sigma, T) \in \mathbb{R} \times \mathbb{S}^n$. The stationarity equations with respect to $C, X, S, \widetilde{y}$, respectively, are

$$T = 0, \qquad \sigma A + \Gamma^Y = 0, \qquad -T - \Gamma^Z = 0, \qquad -\langle \overline{A}, T\rangle = 0.$$

Thus $T = \Gamma^Z = 0$ and $\Gamma^Y = -\sigma A$. The spectral relation gives $\sigma \bar{X} A = 0$. Taking the trace and using $\langle A, \bar{X}\rangle = 1$ yields $\sigma = 0$. Hence all equality and spectral multipliers vanish. □

Proof (Proof of Proposition 12) Fix $\rho > 0$ and the multipliers, and write the equality multiplier as $\mu = (\sigma, T)$. Set $c := L_\rho(x^0)$ and $\delta := \sqrt{2c/\rho}$. For any point in $\mathcal{S}_\rho(x^0)$, the nonnegative objective terms in (20) bound C and X. Set $Q := S + Z$ and $E := C - \widetilde{y}\,\overline{A} - S$. The shifted linking and equality penalties give

$$\|X - Y\|_F \le \delta + \frac{\|M^G\|_F}{\rho}, \qquad \|Q\|_F \le \delta + \frac{\|M^H\|_F}{\rho},$$
$$\|E\|_F \le \delta + \frac{\|T\|_F}{\rho}.$$

Thus Y, Q, and E are bounded throughout the sublevel set. Since $P := -Z \succeq 0$ and $S = Q + P$,

$$R := \widetilde{y}\,\overline{A} + P = C - Q - E$$

is bounded. Let v_+ and v_- be unit eigenvectors of $\overline{A}$ associated with eigenvalues $\alpha_+ > 0$ and $\alpha_- < 0$. Since $P \succeq 0$,

$$\widetilde{y}\alpha_+ \le v_+^\top R v_+, \qquad \widetilde{y}\alpha_- \le v_-^\top R v_-.$$

Choose M so that $\|R\|_2 \le M$ throughout the sublevel set. Then

$$-\frac{M}{|\alpha_-|} \le \widetilde{y} \le \frac{M}{\alpha_+}.$$

Consequently $P = R - \widetilde{y}\,\overline{A}$ is bounded, and hence so are $Z = -P$ and $S = Q + P$. □

These sublevel-set bounds depend on both signs of the spectrum of $\overline{A}$ and need not be uniform across instances or outer iterations. The indefiniteness assumption cannot be discarded from this result. For example, if $0 \ne \overline{A} \succeq 0$, the inner-domain ray

$$C = X = Y = 0, \qquad \widetilde{y} = -t, \qquad S = t\overline{A}, \qquad Z = -t\overline{A}, \qquad t \ge 0,$$

has $(Y, Z) \in \Omega$, zero linking and matrix-equality residuals, and scalar-equality residual -1. For fixed multipliers and ρ, the augmented Lagrangian is constant along this unbounded ray. Hence any sublevel set whose reference point lies on the ray is unbounded.

C.3 Uniform bounds across outer iterations

Proof (Proof of Corollary 5) Lemma 2 gives $L_k(x^{k+1}) \le \bar{L} := \bar{L}_{\mathrm{ref}}$ for every k. Both objectives are nonnegative, so $\bar{L} \ge 0$. Using $\rho_k \ge \rho_0 > 0$ and $\|\bar{\Upsilon}^k\| \le M_{\mathrm{saf}}$ in the squared-distance form of L_k gives

$$\begin{aligned} \operatorname{dist}(\Psi(x^{k+1}), C) &\le \operatorname{dist}\left(\Psi(x^{k+1}) + \frac{\bar{\Upsilon}^k}{\rho_k}, C\right) + \frac{M_{\mathrm{saf}}}{\rho_k} \\ &\le \sqrt{\frac{2\bar{L}}{\rho_k}} + \frac{M_{\mathrm{saf}}}{\rho_k} \le D := \sqrt{\frac{2\bar{L}}{\rho_0}} + \frac{M_{\mathrm{saf}}}{\rho_0}. \end{aligned} \tag{93}$$

Thus every equality residual, linking residual and inequality positive part is uniformly bounded by D, without an upper bound on ρ_k.

For Family I, the objective and the inequality positive part yield

$$\|X^{k+1}\|_F \le \|C_{\mathrm{obs}}\|_F + \sqrt{2\bar{L}}, \qquad \|I - S^{k+1}\|_F^2 \le r + \vartheta + D.$$

Hence X, S are bounded, and the linking residuals bound Y, Z. These estimates apply to the possibly infeasible outer points.

For Family II, the objective bounds C, X. Suppress the superscript $k+1$ and set $E = C - \widetilde{y}\overline{A} - S$, $Q = S + Z$ and $P = -Z \succeq 0$. Then $\|E\|_F, \|Q\|_F \le D$, and

$$R := C - Q - E = \widetilde{y}\overline{A} + P, \qquad \|R\|_2 \le M := \|\widehat{C}\|_F + \sqrt{2\bar{L}} + 2D.$$

For fixed unit eigenvectors v_+, v_- of $\overline{A}$ with eigenvalues $\alpha_+ > 0 > \alpha_-$, positive semidefiniteness of P gives

$$\widetilde{y}\alpha_\pm \le v_\pm^\top R v_\pm, \qquad -\frac{M}{|\alpha_-|} \le \widetilde{y} \le \frac{M}{\alpha_+}.$$

Consequently P, $Z = -P$, $S = Q + P$ and Y are bounded. The constants depend on the fixed instance and need not be uniform as its spectral gaps or b_0 vary.

Boundedness gives an accumulation point. Proposition 4(ii) makes every accumulation point feasible. Theorem 1, the model LICQ results, and Proposition 3 then give M-stationarity, with Proposition 1 transferring it through the unique slack lift. With positive stopping tolerances, any infinite run would have a feasible accumulation point and hence would satisfy Step 1 at a finite iteration by the stopping argument following Theorem 1. □

D A local limitation of the Family I lifting

Consider $n = 3$, $r = 2$, $C_{\text{obs}} = I_3$, and

$$X_0 = \mathbf{1}\mathbf{1}^\top, \qquad S_0 = I_3 - \frac{1}{3}\mathbf{1}\mathbf{1}^\top.$$

Then $X_0, S_0 \succeq 0$, $X_0 S_0 = 0$, and $\|I_3 - S_0\|_F^2 = 1 < r+\vartheta$, so (X_0, S_0) is feasible for (84). The matrix S_0 has two positive eigenvalues. Hence every sufficiently close positive semidefinite matrix S has rank at least two, and $XS = 0$ then implies $\operatorname{rank}(X) \le 1$. A rank-one correlation matrix has the form $ss^\top$ with $s \in \{\pm 1\}^3$; this is a finite set, and X_0 is its only member sufficiently close to X_0. Thus every lifted feasible point sufficiently close to (X_0, S_0) has $X = X_0$ and the same objective value. In particular, (X_0, S_0) is a non-strict local minimizer of the lifted problem.

For the original rank-constrained problem, let

$$X_\theta = \begin{pmatrix} 1 & \cos\theta & \cos\theta \\ \cos\theta & 1 & \cos(2\theta) \\ \cos\theta & \cos(2\theta) & 1 \end{pmatrix}.$$

This is the Gram matrix of the unit vectors $(1,0)^\top$, $(\cos\theta, \sin\theta)^\top$, and $(\cos\theta, -\sin\theta)^\top$. Therefore $X_\theta \succeq 0$, $\operatorname{diag}(X_\theta) = \mathbf{1}$, $\operatorname{rank}(X_\theta) \le 2$, and $X_\theta \to X_0$. For all sufficiently small nonzero θ,

$$\frac{1}{2}\|X_\theta - I_3\|_F^2 = 2\cos^2\theta + \cos^2(2\theta) < 3 = \frac{1}{2}\|X_0 - I_3\|_F^2.$$

Moreover, with $t = \theta^2$ the path $X_{\sqrt{t}}$ is differentiable from the right at $t = 0$, with

$$X_{\sqrt{t}} = X_0 + tH + o(t), \qquad H = \begin{pmatrix} 0 & -\frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & 0 & -2 \\ -\frac{1}{2} & -2 & 0 \end{pmatrix}.$$

Thus H is a Bouligand tangent direction to the original feasible set at X_0, and $\langle X_0 - I_3, H\rangle = -6$. Equivalently,

$$\frac{d}{dt}\frac{1}{2}\left\|X_{\sqrt{t}} - I_3\right\|_F^2\Big|_{t=0+} = -6 < 0.$$

Thus X_0 is neither a local minimizer nor first-order stationary with respect to feasible tangent directions of (83).

Nevertheless, (X_0, S_0) is M-stationary for the lifted formulation. Indeed, the scalar inequality is inactive, and the multipliers

$$\lambda = 0, \qquad \mu = -2\mathbf{1}, \qquad \Gamma^Y = 3S_0, \qquad \Gamma^Z = 0$$

satisfy

$$(X_0 - I_3) + \operatorname{Diag}(\mu) + \Gamma^Y = 0, \qquad 2\lambda(S_0 - I_3) - \Gamma^Z = 0.$$

At $u_0 = (X_0, -S_0)$, the spectrum of $X_0 - S_0$ is $\{3, -1, -1\}$, so the common zero block is empty and $\mathcal{K}_M(u_0) = \mathcal{K}_W(u_0)$. The pair (Γ^Y, Γ^Z) satisfies the corresponding spectral relations: $\Gamma^Y = 3S_0$ has only a $\gamma\gamma$ block, while $\Gamma^Z = 0$. This example shows that a lifted M-stationary local minimizer need not project to a local minimizer or a point stationary with respect to feasible tangent directions of the original rank-constrained problem.

E NP-hardness of exact slack projection

Proof We establish the hardness claim in Remark 4 by reducing Simple Max-Cut [13]. Let G be a simple unweighted graph on $n \ge 2$ vertices with adjacency entries g_{ij}, and denote its maximum cut size by $\operatorname{mc}(G)$. Set

$$d_i = i^2, \qquad D = \operatorname{Diag}(d_1, \dots, d_n), \qquad C_{ij} = g_{ij}\frac{|d_i - d_j|}{ij},$$
$$\varepsilon = (1024n^{16})^{-1}, \qquad A = D + \varepsilon C, \qquad B = -D + \varepsilon C.$$

Throughout, $\|\cdot\| = \|\cdot\|_F$ and $[U, V] = UV - VU$. Put $b = \|C\|$, $H = DC + CD$, and $M = \|H\|$. Then $b \le n^2$, $M \le 2n^4$, and $\varepsilon b < 1$, so $A \succ 0$ and $B \prec 0$.

Let $\mathcal{P} = \{R \in \mathbb{S}^n : R^2 = R\}$. Every complementary pair has orthogonal supports. For fixed $R \in \mathcal{P}$, the minimizing supported pair is $Y = RAR$, $Z = (I - R)B(I - R)$. Writing $v(A, B)$ for the value of (57), expansion therefore gives

$$v(A, B) = c_0 - \max_{R\in\mathcal{P}} h(R), \qquad c_0 = \|D\|^2 + \varepsilon^2 b^2,$$
$$h(R) = -\|[D, R]\|^2 + 2\varepsilon\langle H, R\rangle - \varepsilon^2\|[C, R]\|^2.$$

For a cut S, set $R_S = \operatorname{Diag}(\mathbf{1}_S)$ and $Q_S = I - R_S$. On symmetric matrices $T = R_S T Q_S + Q_S T R_S$, define

$$q_S(T) = -\|[D, T]\|^2 + 2\varepsilon\langle H, T\rangle - \varepsilon^2\|[C, R_S]\|^2.$$

Let T_S have cross-block entries $(T_S)_{ij} = \varepsilon H_{ij}/(d_i - d_j)^2$ for $i \in S$, $j \notin S$, with symmetric counterparts and zero same-side blocks. Completing squares yields

$$q_S(T) = 8\varepsilon^2 \operatorname{cut}_G(S) - \|[D, T - T_S]\|^2, \qquad \|T_S\| \le \varepsilon M.$$

Take a global maximizer R of h, which exists by compactness of $\mathcal{P}$. Since H has zero diagonal and $|d_i - d_j| \ge 1$ for $i \ne j$,

$$0 \le h(R) \le -t^2 + 2\varepsilon M t, \qquad t := \|\operatorname{offdiag} R\| \le 2\varepsilon M \le \tfrac{1}{2}.$$

Round the diagonal of R to obtain R_S and write $\Delta = R - R_S = T + W$, where T is the cross-block part. The identities $R_{ii}(1 - R_{ii}) = \sum_{j \neq i} R_{ij}^2$ and $R_S\Delta + \Delta R_S + \Delta^2 = \Delta$ imply $\|\Delta\| \leq 2t$ and $\|W\| \leq \|\Delta\|^2 \leq 4t^2$. The matrices $[D, T]$ and $[D, W]$ have disjoint block supports. Using $\|[C, R_S]\| \leq 2b$ and $\|[C, \Delta]\| \leq 4bt$, expansion gives

$$\begin{aligned} h(R) &\leq q_S(T) + 8\varepsilon M t^2 + 24\varepsilon^2 b^2 t \\ &\leq 8\varepsilon^2 \operatorname{mc}(G) + 32\varepsilon^3 M^3 + 48\varepsilon^3 b^2 M \\ &\leq 8\varepsilon^2 \operatorname{mc}(G) + 352\varepsilon^3 n^{12}. \end{aligned}$$

For the reverse bound, fix any cut S and set $K = T_S R_S - R_S T_S$ and $\widetilde{R} = e^K R_S e^{-K}$. Then $K^\top = -K$, $[K, R_S] = T_S$, and $q := \|K\| = \|T_S\| \leq \varepsilon M \leq \frac{1}{4}$. The first and second derivatives of $e^{sK} R_S e^{-sK}$ have norms at most q and $2q^2$. Thus, for $\widetilde{\Delta} = \widetilde{R} - R_S$,

$$\|\widetilde{\Delta}\| \leq q, \qquad \|\widetilde{\Delta} - T_S\| \leq q^2.$$

Expanding h and using these bounds yields

$$\begin{aligned} h(\widetilde{R}) &\geq q_S(T_S) - 9n^4 q^3 - 2\varepsilon M q^2 - 9\varepsilon^2 b^2 q \\ &\geq 8\varepsilon^2 \operatorname{cut}_G(S) - 106\varepsilon^3 n^{16}. \end{aligned}$$

Taking a maximum cut and combining both bounds gives

$$\left| v(A, B) - c_0 + 8\varepsilon^2 \operatorname{mc}(G) \right| \leq 512\varepsilon^3 n^{16} = \frac{\varepsilon^2}{2}.$$

Hence, for every integer k,

$$\operatorname{mc}(G) \geq k \quad \Longleftrightarrow \quad v(A, B) \leq c_0 - 8\varepsilon^2 \left(k - \tfrac{1}{2}\right).$$

The matrices and threshold are rational, have polynomial encoding length, and can be constructed in polynomial time. This completes the reduction. □

References

1. Absil, P.A., Mahony, R., Sepulchre, R.: Optimization Algorithms on Matrix Manifolds. Princeton University Press, Princeton, NJ (2008)
2. Andreani, R., Haeser, G., Secchin, L.D., Silva, P.J.S.: New sequential optimality conditions for mathematical programs with complementarity constraints and algorithmic consequences. SIAM Journal on Optimization **29**(4), 3201–3230 (2019). DOI 10.1137/18M121040X
3. Andreani, R., Secchin, L.D., Silva, P.J.S.: Convergence properties of a second order augmented Lagrangian method for mathematical programs with complementarity constraints. SIAM Journal on Optimization **28**(3), 2574–2600 (2018). DOI 10.1137/17M1125698
4. Anitescu, M., Tseng, P., Wright, S.J.: Elastic-mode algorithms for mathematical programs with equilibrium constraints: Global convergence and stationarity properties. Mathematical Programming **110**, 337–371 (2007). DOI 10.1007/s10107-006-0005-4
5. Aswani, A., Shen, Z.J., Siddiq, A.: Inverse optimization with noisy data. Operations Research **66**(3), 870–892 (2018). DOI 10.1287/opre.2017.1705
6. Bi, S., Han, L., Pan, S.: Approximation of rank function and its application to the nearest low-rank correlation matrix. Journal of Global Optimization **57**, 1113–1137 (2013). DOI 10.1007/s10898-012-0007-0
7. Brigo, D., Mercurio, F.: Calibrating LIBOR. Risk Magazine **15**, 117–122 (2002)
8. Chan, T.C.Y., Mahmood, R., Zhu, I.Y.: Inverse optimization: Theory and applications. Operations Research **73**(2), 1046–1074 (2025). DOI 10.1287/opre.2022.0382
9. Ding, C., Sun, D., Ye, J.J.: First order optimality conditions for mathematical programs with semidefinite cone complementarity constraints. Mathematical Programming **147**(1–2), 539–579 (2014). DOI 10.1007/s10107-013-0735-z. URL `https://doi.org/10.1007/s10107-013-0735-z`
10. Edelman, A., Arias, T.A., Smith, S.T.: The geometry of algorithms with orthogonality constraints. SIAM Journal on Matrix Analysis and Applications **20**(2), 303–353 (1998). DOI 10.1137/S0895479895290954
11. Fan, S., Pagliardini, M., Jaggi, M.: DOGE: Domain reweighting with generalization estimation. In: Proceedings of the 41st International Conference on Machine Learning, *Proceedings of Machine Learning Research*, vol. 235, pp. 12895–12915. PMLR (2024). URL `https://proceedings.mlr.press/v235/fan24e.html`
12. Franceschi, L., Frasconi, P., Salzo, S., Grazzi, R., Pontil, M.: Bilevel programming for hyperparameter optimization and meta-learning. In: Proceedings of the 35th International Conference on Machine Learning, *Proceedings of Machine Learning Research*, vol. 80, pp. 1568–1577. PMLR (2018). URL `https://proceedings.mlr.press/v80/franceschi18a.html`
13. Garey, M.R., Johnson, D.S., Stockmeyer, L.: Some simplified NP-complete graph problems. Theoretical Computer Science **1**(3), 237–267 (1976). DOI 10.1016/0304-3975(76)90059-1
14. Ghatrani, Z., Ghate, A.: Inverse optimization in semi-definite programs to impute unknown constraint matrices. Computers & Operations Research **168**, 106681 (2024). DOI 10.1016/j.cor.2024.106681
15. Guo, L., Deng, Z.: A new augmented Lagrangian method for MPCCs—theoretical and numerical comparison with existing augmented Lagrangian methods. Mathematics of Operations Research **47**(2), 1229–1246 (2022). DOI 10.1287/moor.2021.1165
16. Hu, X., Ralph, D.: Convergence of a penalty method for mathematical programming with complementarity constraints. Journal of Optimization Theory and Applications **123**, 365–390 (2004). DOI 10.1007/s10957-004-5154-0
17. Huber, P.J.: Robust estimation of a location parameter. The Annals of Mathematical Statistics **35**(1), 73–101 (1964). DOI 10.1214/aoms/1177703732
18. Iyengar, G., Kang, W.: Inverse conic programming with applications. Operations Research Letters **33**(3), 319–330 (2005). DOI 10.1016/j.orl.2004.04.007

19. Izmailov, A.F., Solodov, M.V., Uskov, E.I.: Global convergence of augmented Lagrangian methods applied to optimization problems with degenerate constraints, including problems with complementarity constraints. SIAM Journal on Optimization **22**(4), 1579–1606 (2012). DOI 10.1137/120868359
20. Jia, X., Kanzow, C., Mehlitz, P., Wachsmuth, G.: An augmented Lagrangian method for optimization problems with structured geometric constraints. Mathematical Programming **199**, 1365–1415 (2023). DOI 10.1007/s10107-022-01870-z
21. Kadrani, A., Dussault, J.P., Benchakroun, A.: A new regularization scheme for mathematical programs with complementarity constraints. SIAM Journal on Optimization **20**(1), 78–103 (2009). DOI 10.1137/070705490
22. Kanzow, C., Schwartz, A.: A new regularization method for mathematical programs with complementarity constraints with strong convergence properties. SIAM Journal on Optimization **23**(2), 770–798 (2013). DOI 10.1137/100802487
23. Li, Q., Qi, H.: A sequential semismooth Newton method for the nearest low-rank correlation matrix problem. SIAM Journal on Optimization **21**(4), 1641–1666 (2011). DOI 10.1137/090771181
24. Liang, Y.C., Zhu, C.Y., Zhang, S.J., Lin, G.H., Zhu, X.: Augmented lagrangian method for mathematical programs with second-order cone complementarity constraints (2026). URL `https://arxiv.org/abs/2606.16820v1`. Version 1, June 15, 2026
25. Lin, G.H., Fukushima, M.: Some exact penalty results for nonlinear programs and mathematical programs with equilibrium constraints. Journal of Optimization Theory and Applications **118**, 67–80 (2003). DOI 10.1023/A:1024787424532
26. Liu, D.C., Nocedal, J.: On the limited memory BFGS method for large scale optimization. Mathematical Programming **45**(1–3), 503–528 (1989). DOI 10.1007/BF01589116
27. Luo, Z.Q., Pang, J.S., Ralph, D.: Mathematical Programs with Equilibrium Constraints. Cambridge University Press, Cambridge (1996)
28. Madariaga, J.I., Ramírez, H.: Facial approach for constructing stationary points for mathematical programs with cone complementarity constraints. Journal of Optimization Theory and Applications **204**, 15 (2025). DOI 10.1007/s10957-024-02562-8
29. Mehlitz, P.: Asymptotic stationarity and regularity for nonsmooth optimization problems. Journal of Nonsmooth Analysis and Optimization **1**, 6575 (2020). DOI 10.46298/jnsao-2020-6575
30. Ralph, D., Wright, S.J.: Some properties of regularization and penalization schemes for MPECs. Optimization Methods and Software **19**(5), 527–556 (2004). DOI 10.1080/10556780410001709439
31. Rockafellar, R.T., Wets, R.J.B.: Variational Analysis, *Grundlehren der mathematischen Wissenschaften*, vol. 317. Springer, Berlin (1998). DOI 10.1007/978-3-642-02431-3
32. Santos, D.O., Costa, C.M., Krulikovski, E.H.M., Geremia, M.: Augmented Lagrangian methods for nonlinear semidefinite programming with complementarity constraints (2026). ArXiv:2606.08424
33. Scheel, H., Scholtes, S.: Mathematical programs with complementarity constraints: Stationarity, optimality, and sensitivity. Mathematics of Operations Research **25**(1), 1–22 (2000). DOI 10.1287/moor.25.1.1.15213
34. Scholtes, S.: Convergence properties of a regularization scheme for mathematical programs with complementarity constraints. SIAM Journal on Optimization **11**(4), 918–936 (2001). DOI 10.1137/S1052623499361233
35. Wu, J., Zhang, L.: On properties of the bilinear penalty function method for mathematical programs with semidefinite cone complementarity constraints. Set-Valued and Variational Analysis **23**, 277–294 (2015). DOI 10.1007/s11228-014-0295-2
36. Wu, J., Zhang, L., Zhang, Y.: Mathematical programs with semidefinite cone complementarity constraints: Constraint qualifications and optimality conditions. Set-Valued and Variational Analysis **22**, 155–187 (2014). DOI 10.1007/s11228-013-0242-7
37. Wu, J., Zhang, Y., Zhang, L., Lu, Y.: A sequential convex program approach to an inverse linear semidefinite programming problem. Asia-Pacific Journal of Operational Research **33**(4), 1650025 (2016). DOI 10.1142/S0217595916500251
38. Wu, L.: Fast at-the-money calibration of the LIBOR market model using Lagrange multipliers. Journal of Computational Finance **6**, 39–77 (2003)
39. Yan, T., Fukushima, M.: Smoothing method for mathematical programs with symmetric cone complementarity constraints. Optimization **60**(1–2), 113–128 (2011). DOI 10.1080/02331934.2010.541458
40. Ye, J.J., Zhou, J.: First-order optimality conditions for mathematical programs with second-order cone complementarity constraints. SIAM Journal on Optimization **26**(4), 2820–2846 (2016). DOI 10.1137/16M1055554
41. Zhang, Z., Wu, L.: Optimal low-rank approximation to a correlation matrix. Linear Algebra and Its Applications **364**, 161–187 (2003)