First-order methods for semidefinite programming

Rank identification and local linear convergence

Xin Jiang

Cornell ORIE $\Rightarrow$ U. of Houston ISE

joint work with Shucheng Kang (Harvard) and Heng Yang (Harvard)

International Conference on Continuous Optimization (ICCOPT)

July 22, 2025

Semidefinite programs (SDP)

primal: minimize $\langle C, X \rangle$
subject to $\mathcal{A}(X) = b$
$X \in \mathbb{S}^n_+$

dual: maximize $\langle b, y \rangle$
subject to $\mathcal{A}^*(y) + S = C$
$S \in \mathbb{S}^n_+$

$\mathcal{A}\colon \mathbb{S}^n \to \mathbb{R}^m$ is a linear mapping, and $\mathcal{A}^*$ is its adjoint

$$\mathcal{A}(X) = \big(\langle A_1, X \rangle, \langle A_2, X \rangle, \ldots, \langle A_m, X \rangle\big)$$

Interior-point methods

- general-purpose implementations for dense problems do not scale well
- each iteration involves computations with complexity m^3, $m^2 n^2$, mn^3
- customization to exploit problem structure is difficult

First-order proximal splitting methods

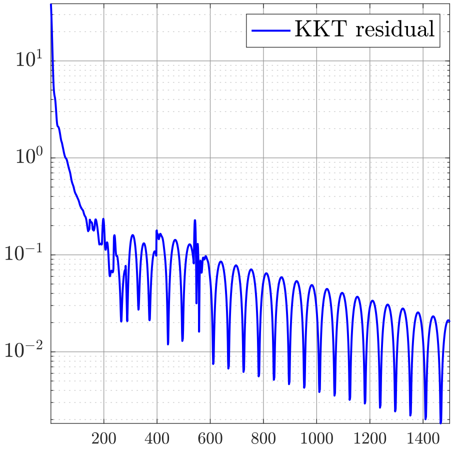
- for example, ADMM, Douglas–Rachford splitting (DRS), primal–dual hybrid gradient (PDHG)
- exploit structure in linear constraints is straightforward
- require eigenvalue decompositions for projections on positive semidefinite (PDS) cones

ADMM for (dual) SDP

$$y_{k+1} = (\mathcal{A}\mathcal{A}^*)^{-1}\left(\sigma^{-1}b - \mathcal{A}(\sigma^{-1}X_k + S_k - C)\right)$$
$$S_{k+1} = \Pi_{\mathbb{S}^n_+}\left(C - \mathcal{A}^* y_{k+1} - \sigma^{-1}X_k\right)$$
$$X_{k+1} = X_k + \sigma\left(S_{k+1} + \mathcal{A}^* y_{k+1} - C\right)$$

- $(\mathcal{A}\mathcal{A}^*)^{-1}$ involves one factorization or solving a linear system per iteration
- $\Pi_{\mathbb{S}^n_+}$ requires an eigenvalue decomposition

- solves SDPs to moderate accuracy
- suffers from slow sublinear worst-case rate

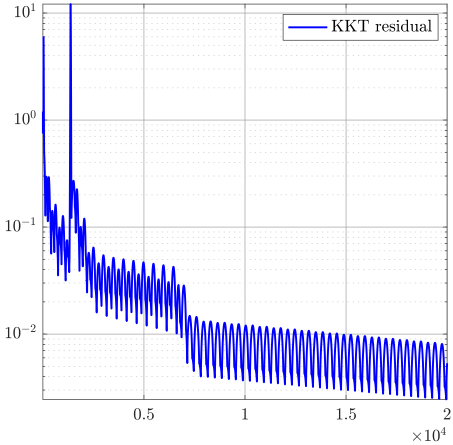


Wen, Goldfarb & Yin (2010)

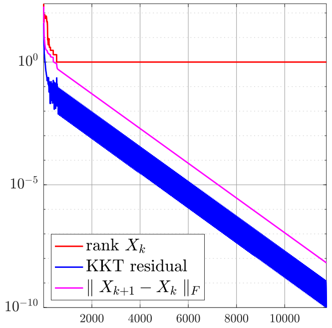
PDHG for SDP

$$X_{k+1} = \Pi_{\mathbb{S}^n_+}\big(X_k - \tau(C - \mathcal{A}^* y_k)\big)$$
$$y_{k+1} = y_k - \sigma\big(\mathcal{A}(2X_{k+1} - X_k) - b\big)$$

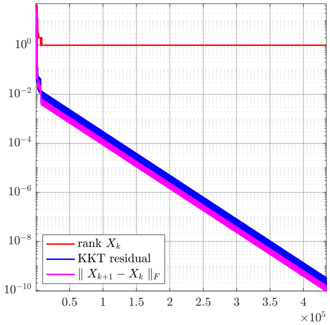
- $\Pi_{\mathbb{S}^n_+}$ requires an eigenvalue decomposition
- only requires linear mappings $\mathcal{A}$ and $\mathcal{A}^*$



First-order methods for SDP



ADMM



PDHG

- **Local (R-)linear convergence**
- **Rank identification:** after finitely many iterations, X_k finds and maintains the solution rank

$$\operatorname{rank}(X_k) = \operatorname{rank}(X_\star) \quad \text{for } k \geq \bar{k}_{\text{ID}}$$

Outline

Rank identification

Linear convergence

Open questions and future directions

One-step ADMM (for SDP)

primal–dual solutions are simultaneously diagonalizable and assume strict complementarity

$$X_\star = Q_\star \begin{bmatrix} \Lambda_X & 0 \\ 0 & 0 \end{bmatrix} Q_\star^\mathsf{T} \in \mathbb{S}_+^n, \quad S_\star = Q_\star \begin{bmatrix} 0 & 0 \\ 0 & \Lambda_S \end{bmatrix} Q_\star^\mathsf{T} \in \mathbb{S}_+^n, \quad Z_\star = X_\star - S_\star = Q_\star \begin{bmatrix} \Lambda_X & 0 \\ 0 & -\Lambda_S \end{bmatrix} Q_\star^\mathsf{T},$$

where $Q_\star$ is orthogonal, $\lambda_1 \geq \cdots \geq \lambda_r > 0 > \lambda_{r+1} \geq \cdots \geq \lambda_n$, and

$$\Lambda_X := \operatorname{diag}(\lambda_1, \ldots, \lambda_r), \qquad \Lambda_S := -\operatorname{diag}(\lambda_{r+1}, \ldots, \lambda_n),$$

One-step ADMM (take $\sigma = 1$ for simplicity)

$$Z_{k+1} = \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}\mathcal{A}(-2\Pi_{\mathbb{S}_+^n}(Z_k) + Z_k) + \Pi_{\mathbb{S}_+^n}(Z_k) + \mathcal{A}^*(\mathcal{A}\mathcal{A}^*)^{-1}(\mathcal{A}C + b) - C,$$

- key observation: X_k and S_k share the same eigenspace; so $Z_k = X_k - S_k$
- from Z_k, we can extract X_k and S_k: $X_k = \Pi_{\mathbb{S}_+^n}(Z_k)$ and $S_k = \Pi_{\mathbb{S}_+^n}(-Z_k)$

A direct proof of rank identification

Two equivalent statements assume ADMM converges to a strictly complementary solution:

$$\operatorname{rank} X_\star + \operatorname{rank} S_\star = n$$

- there exists $\bar{k}_{\mathrm{ID}} \in \mathbb{N}$ such that for any integer $k \geq \bar{k}_{\mathrm{ID}}$, it holds that

$$\operatorname{rank} X_k = \operatorname{rank} X_\star =: r, \qquad \operatorname{rank} S_k = \operatorname{rank} S_\star = n - r$$

- if $\|Z_k - Z_\star\|_2 \leq \min\{\lambda_r, -\lambda_{r+1}\}$ (recall $\{\lambda_i\}$ are the eigenvalues of $Z_\star$), then

$$\gamma_r := \operatorname{eig}_r(Z_k) > 0, \qquad \gamma_{r+1} := \operatorname{eig}_{r+1}(Z_k) < 0,$$

where γ_r and γ_{r+1} the rth and $(r+1)$st largest eigenvalue of Z_k, respectively

Proof: from Weyl's inequality, we see

$$\gamma_r \geq \lambda_r - \|Z_k - Z_\star\|_2 > \lambda_r - \min\{\lambda_r, -\lambda_{r+1}\} \geq 0$$
$$\gamma_{r+1} \leq \lambda_{r+1} + \|Z_k - Z_\star\|_2 < \lambda_{r+1} + \min\{\lambda_r, -\lambda_{r+1}\} \leq 0$$

so $\operatorname{rank} X_k = \operatorname{rank}\left(\Pi_{\mathbb{S}^n_+}(Z_k)\right) = r = \operatorname{rank}\left(\Pi_{\mathbb{S}^n_+}(Z_\star)\right) = \operatorname{rank} X_\star$

Zoom out: partial smoothness and activity identification

such identification is not unique in ADMM and/or SDP

Partly smooth function: $f : \mathbb{E} \to \overline{\mathbb{R}}$ is partly smooth at x relative to a manifold $\mathcal{M}$ if

- restricted smoothness: the restriction $f|_{\mathcal{M}}$ is smooth around x
- normal sharpness: $f'(x; -v) + f'(x; v) > 0$ for all nonzero directions v in $\mathcal{N}_{\mathcal{M}}(x)$

together with mild conditions on its subdifferential ∂f

Activity identification

- suppose f is partly smooth at $x_\star \in \operatorname{argmin} f$ w.r.t. $\mathcal{M}$ and $s_\star \in \operatorname{relint} \partial f(x_\star)$
- suppose the sequence (x_k, s_k) satisfies $s_k \in \partial f(x_k)$ and converges to $(x_\star, s_\star)$
- for sufficiently large k, it holds that $x_k \in \mathcal{M}$

Lewis (2002); Drusvyatskiy & Lewis (2014); Fadili, Malick & Feyré (2018)

Examples and counter-examples

- **Piecewise linear function:** $f(x) = \max_{i \in \mathcal{I}}\{\langle a_i, x\rangle + b_i\}$; the identifiable set at $\bar{x}$ is

$$\mathcal{M}_{\bar{x}} = \{x \mid \mathcal{I}(x) = \mathcal{I}(\bar{x})\}, \quad \text{where } \mathcal{I}(x) = \{i \in \mathcal{I} \mid \langle a_i, x\rangle + b_i = f(x)\}$$

- **Indicator of PSD cone:** consider $\overline{X} \in \mathbb{S}^n_+$ and there exists $\overline{S} \in \operatorname{relint} \mathcal{N}_{\mathbb{S}^n_+}(\overline{X})$

$$\mathcal{M}_{\overline{X}} = \{X \in \mathbb{S}^n_+ \mid \operatorname{rank} X = \operatorname{rank} \overline{X}\}$$

  - the regularity condition $\overline{S} \in \operatorname{relint} \mathcal{N}_{\mathbb{S}^n_+}(\overline{X})$ amounts to strict complementarity in SDP
  - without strict complementarity, PSD cone may not admit an identifiable set

- **Counter-example:** $f(x, y) = \sqrt{x^4 + y^2}$ is not partly smooth and does not admit a manifold

Lewis (2002); Drusvyatskiy & Lewis (2014); Fadili, Malick & Feyré (2018)

Proximal (splitting) methods for SDP

primal: minimize $\langle C, X\rangle$ subject to $\mathcal{A}(X) = b$, $X \in \mathbb{S}^n_+$

dual: maximize $\langle b, y\rangle$ subject to $\mathcal{A}^*(y) + S = C$, $S \in \mathbb{S}^n_+$

- proximal splitting methods: ADMM for dual SDP is DRS applied to primal SDP

$$\text{minimize} \quad \delta_{\mathbb{S}^n_+}(X) + \big(\langle C, X\rangle + \delta_{\{X \mid \mathcal{A}(X)=b\}}(X)\big)$$

  - $\delta_{\mathbb{S}^n_+}$ is partly smooth at $X_\star$ with respect to the fixed-rank manifold (under SC)
  - (X_k, S_k) satisfies $S_k \in \mathcal{N}_{\mathbb{S}^n_+}(X_k)$ and converges to $(X_\star, S_\star)$
  - so the X_k iterates identify the solution rank for sufficiently large k

- augmented Lagrangian method: PPM applied to the dual $h(y) = \langle b, y\rangle + \delta_{\mathbb{S}^n_+}(C - \mathcal{A}^*(y))$

$$X_{k+1} = \operatorname{argmin}\{\langle C, X\rangle + \tfrac{\rho}{2}\|\mathcal{A}X - b + \tfrac{1}{\rho}y_k\|_2^2 \mid X \succeq 0\}$$

$$y_{k+1} = y_k + \rho(\mathcal{A}X_{k+1} - b)$$

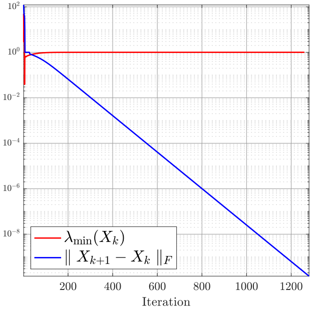
  - h may not be partly smooth at $y_\star$
  - additional condition is needed, *e.g.*, primal solution is unique [DLY25]

Numerical evidence

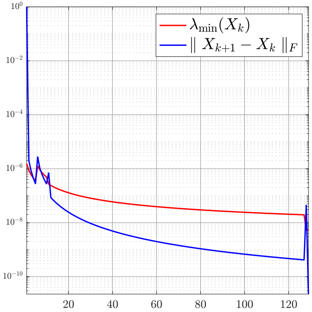
- apply ADMM and augmented Lagrangian method (ALM) to the SDP reformulation of

$$\text{minimize} \quad f(x,y) := \sqrt{x^4 + y^2}$$

- in the reformulated SDP, $\operatorname{rank}(X_\star) = 1$ and $\lambda_1(X_\star) = 1$
- f is not partly smooth at 0, so ALM does not have rank identification, whereas ADMM does



ADMM



ALM

Outline

Rank identification

Linear convergence

Open questions and future directions

A refined error bound for PSD cone projection

for a nonsingular $Z \in \mathbb{S}^n$, denote its eigenvalue decomposition by

$$Z = Q\,\mathrm{diag}(\lambda_1, \dots, \lambda_r, \lambda_{r+1}, \dots, \lambda_n) Q^\top = \begin{bmatrix} Q_X & Q_S \end{bmatrix} \begin{bmatrix} \Lambda_X & 0 \\ 0 & \Lambda_S \end{bmatrix} \begin{bmatrix} Q_X^\top \\ Q_S^\top \end{bmatrix}$$

where $\lambda_1 \geq \cdots \geq \lambda_r > 0 > \lambda_{r+1} \geq \cdots \geq \lambda_n$

- previous result [SS02]: $\left\| \Pi_{\mathbb{S}^n_+}(Z + \Delta) - \Pi_{\mathbb{S}^n_+}(Z) - \left(\Pi_{\mathbb{S}^n_+}(Z)\right)'(\Delta) \right\|_2 \lesssim \|\Delta\|_2^2$
- when $Q = I$, for all $\Delta \in \mathbb{S}^n$ with norm sufficiently small, it holds that

$$\left\| \Pi_{\mathbb{S}^n_+}(Z + \Delta) - \Pi_{\mathbb{S}^n_+}(Z) - \left(\Pi_{\mathbb{S}^n_+}(Z)\right)'(\Delta) \right\|_2 \lesssim \|\Delta_O\|_2 \|\Delta\|_2, \quad \text{where } \Delta = \begin{bmatrix} \Delta_X & \Delta_O^\top \\ \Delta_O & \Delta_S \end{bmatrix}$$

- in general: $\left\| \Pi_{\mathbb{S}^n_+}(Z + \Delta) - \Pi_{\mathbb{S}^n_+}(Z) - \left(\Pi_{\mathbb{S}^n_+}(Z)\right)'(\Delta) \right\|_2 \lesssim \|Q_S^\top \Delta_O Q_X\|_2 \|\Delta\|_2$

Local linear convergence with nondegeneracy

Local linearization of ADMM one-step ADMM can be further reformulated as

$$Z_{k+1} - Z_\star = \mathcal{F}(Z_k - Z_\star) + \Psi_k,$$

where $\mathcal{F}\colon \mathbb{S}^n \to \mathbb{S}^n$ is a linear, firmly nonexpansive with $\|\mathcal{F} - \Pi_{\mathrm{Fix}(\mathcal{F})}\|_{\mathrm{op}} < 1$, and

$$\|\Psi_k\|_{\mathsf{F}} \lesssim \|\Delta_{k,O}\|_{\mathsf{F}} \|\Delta_k\|_{\mathsf{F}}, \quad \text{when } \Delta_k := Z_k - Z_\star \text{ is sufficiently small}$$

Local linear convergence with nondegeneracy

primal nondegeneracy: $\mathcal{N}_{X_\star} \cap \mathrm{Range}(\mathcal{A}^*) = \{0\}$, dual nondegeneracy: $\mathcal{N}_{S_\star} \cap \mathrm{Null}(\mathcal{A}) = \{0\}$

- under strict complementarity (SC), nondegeneracy implies uniqueness of primal–dual solutions
- in this case, $\mathrm{Fix}(\mathcal{F}) = \{0\}$ and

$$\|Z_{k+1} - Z_\star\|_{\mathsf{F}} \le \rho \|Z_k - Z_\star\|_{\mathsf{F}}, \quad \text{for sufficiently large } k \text{ and for any } \rho \in (\|\mathcal{F}\|_{\mathrm{op}}, 1)$$

- same proof holds in non-SC case, recovering [HSZ18] (w/o metric subregularity of KKT operator)

Local (R-)linear convergence without nondegeneracy (ND)

- without ND, $\mathrm{Fix}(\mathcal{F}) \neq \{0\}$ and the above technique can only establish (R-)linear conv. of

$$(\mathrm{Id} - \Pi_{\mathrm{Fix}(\mathcal{F})})\Delta_k, \qquad \Pi_{\mathcal{T}_{S_\star}}(X_k), \qquad \Pi_{\mathcal{T}_{X_\star}}(S_k)$$

  the last two terms are the part of X_k (or S_k) that lies outside the minimal face of $X_\star$ (or $S_\star$)

- consider an affine space $\mathcal{V} := \{X \mid \mathcal{A}X = b\}$

$$\mathrm{dist}(x, \mathcal{V} \cap \mathbb{R}^n_+) \lesssim \mathrm{dist}(x, \mathcal{V}) + [-x]_+ \qquad \text{sharpness}$$

$$\mathrm{dist}(X, \mathcal{V} \cap \mathbb{S}^n_+) \lesssim \left(\mathrm{dist}(X, \mathcal{V}) + [-\lambda_{\min}(X)]_+\right)^{1/2}$$

$$\mathrm{dist}(X, \mathcal{V} \cap \mathbb{S}^n_+ \cap \mathcal{T}^\perp_{S_\star}) \lesssim \mathrm{dist}(X, \mathcal{V}) + [-\lambda_{\min}(X)]_+ + \|\Pi_{\mathcal{T}_{S_\star}}(X)\|_\mathsf{F} \qquad \text{[Sturm00]}$$

- this gives a linear growth condition on the distance to optimality:

$$\mathrm{dist}(Z_k, \mathcal{Z}_\star) \lesssim \|Z_{k+1} - Z_k\|_\mathsf{F} + \|\Pi_{\mathcal{T}_{S_\star}}(X_k)\|_\mathsf{F} + \|\Pi_{\mathcal{T}_{X_\star}}(S_k)\|_\mathsf{F}$$

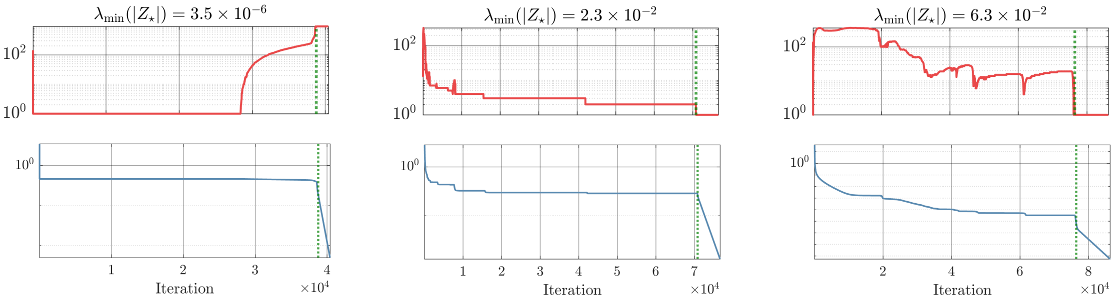
Outline

Rank identification

Linear convergence

Open questions and future directions

Rank identification and linear convergence



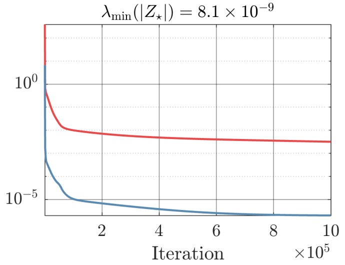
`hamming-11-2` `BQP-r1-30-3` `Quasar-200`

(red curve plot $\mathrm{rank}(X_k)$ and blue curve plots $\|Z_{k+1} - Z_k\|_\mathsf{F}$)

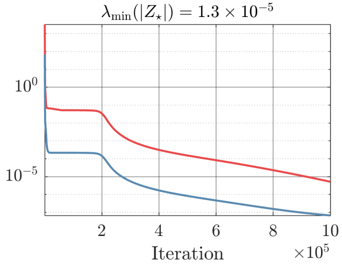
Open questions

- in what type of SDPs is rank identification a necessary condition for (R-)linear convergence
- under which conditions will rank identification and (R-)linear convergence occur simultaneously?

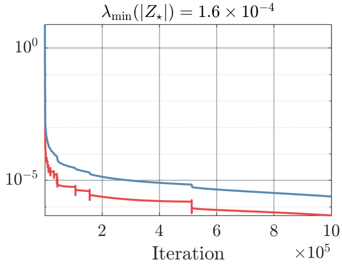
Failure cases



`cnhil10`



`neosfbr25`



`MAXCUT-G11`

- red curve plots KKT residual $r_{\max}$ and blue curve plots $\|Z_{k+1} - Z_k\|_{\mathsf{F}}$
- ADMM fails to achieve $r_{\max} \leq 10^{-10}$ within the budget, with no evident linear convergence
- common feature: min. eigen-val of $Z_\star$ (in abs. val.) is small $(10^{-4} \sim 10^{-9})$, but not exactly zero
- this observation aligns with recent findings in PDHG for LP [LY24]

Conclusion

- **Algorithmic contribution** on ADMM for SDP
  - mild assumption: ADMM converges to a strictly complementary solution
  - rank identification: ADMM identifies the solution rank in finitely many iterations
  - local (R-)linear convergence: a refined error bound for PSD cone projection
- **Empirical contribution**
  - numerical results show rank identification and linear convergence across diverse SDPs
  - demonstrate failure cases linked to near-violations of strict complementarity

Shucheng Kang, Xin Jiang, and Heng Yang.
Local linear convergence of ADMM for SDP under strict complementarity.
arXiv:2503.20142