Multi-stage convex relaxation approach for low-rank structured PSD matrix recovery

Department of Mathematics & Risk Management Institute
National University of Singapore

(Based on a joint work with Shujun Bi and Shaohua Pan)

May 22, 2014

Outline

1. Problem formulation
2. Multi-stage convex relaxation approach
   - Exact penalty for the equivalent MPSDEC
   - A unified framework for multi-stage convex relaxation
3. The power of multi-stage convex relaxation approach
   - General error bound
   - Error bound comparison for the first two stages
   - Error bound comparison for the first k stages
4. Numerical results
5. Conclusion

Outline

1. Problem formulation
2. Multi-stage convex relaxation approach
   - Exact penalty for the equivalent MPSDEC
   - A unified framework for multi-stage convex relaxation
3. The power of multi-stage convex relaxation approach
   - General error bound
   - Error bound comparison for the first two stages
   - Error bound comparison for the first k stages
4. Numerical results
5. Conclusion

Matrix recovery

Given some measurements for a (structured) matrix $\overline{X} \in \mathcal{H}_+^n$ (Hermitian positive semi-definite), how to recover this unknown matrix?

$$b_k = \langle \Theta_k, \overline{X} \rangle + \xi_k, \quad k = 1, 2, \ldots, m.$$

"b_k": observed data (with / without noise) "ξ_k": additive noise

- Possible? — Yes, with the low-rank structure.
- Entries = basis coefficients: $\{e_i e_j^T \mid 1 \leq i \leq n_1, 1 \leq j \leq n_2\}$.

Correlation matrix recovery

A low-rank correlation matrix $\overline{X} \in \mathcal{S}^n_+$:

$$\text{symmetric positive semidefinite} \;\; \text{and} \;\; \text{diag}(\overline{X}) = e.$$

Observations (with / without noise):

$$b_k = \langle \Theta_k, \overline{X} \rangle + \xi_k, \;\; k = 1, \ldots, m.$$

How to recover the unknown correlation matrix $\overline{X}$?

- All diagonal entries are fixed.
- Some off-diagonal entries may also be fixed, [e.g., the correlations among pegged currencies].
- Entries $\rightleftharpoons$ basis coefficients $\{e_i e_i^T | 1 \le i \le n\} \bigcup \{\frac{1}{\sqrt{2}}(e_i e_j^T + e_j e_i^T) | 1 \le i < j \le n\}$.

A low-rank density matrix $\overline{X} \in \mathcal{H}^n_+$ with $n = 2^l$:

$$\text{Hermitian positive semidefinite and } \operatorname{Tr}(\overline{X}) = 1.$$

Observations – Pauli measurements:

$$\langle \Theta_k, \overline{X} \rangle = \operatorname{Re}(\operatorname{Tr}(\Theta_k \overline{X})), \quad k = 1, \ldots, m,$$

where $\Theta_k \in$ Pauli basis $\left\{ \sigma_{s_1} \otimes \cdots \otimes \sigma_{s_l} \mid (s_1, \cdots, s_l) \in \{0, 1, 2, 3\}^l \right\}$ with

$$\sigma_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \ \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \ \sigma_2 = \begin{pmatrix} 0 & -\sqrt{-1} \\ \sqrt{-1} & 0 \end{pmatrix}, \ \sigma_3 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

How to recover the unknown density matrix $\overline{X}$?

Observation model

In this talk, we consider the problem of low-rank structured PSD matrix recovery with noise. The observation model takes the following form

$$b_k = \langle \Theta_k, \overline{X} \rangle + \xi_k, \qquad k = 1, \ldots, m.$$

Define $\mathcal{A}(X) := \big(\langle \Theta_1, X \rangle, \cdots, \langle \Theta_m, X \rangle\big)^T \in \mathbb{R}^m$. Then, we have that

$$b = \mathcal{A}(\overline{X}) + \xi.$$

- Let $\overline{X} \in \mathcal{H}^n_+$ be the unknown low-rank structured true matrix.
- There exists a constant $\delta \geq 0$, $\|\mathcal{A}(\overline{X}) - b\| \leq \delta$.

The rank minimization estimator:

$$\widehat{X} \in \arg\min_{\Omega \cap \mathcal{H}^n_+} \left\{ \mathrm{rank}(X) : \ \|\mathcal{A}(X) - b\| \leq \delta \right\}$$

where Ω is a closed convex set characterizing the structure of $\overline{X}$.

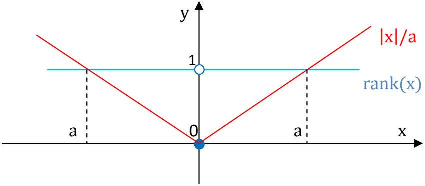
- This is the best possible estimator. But it is very challenging to get the global solution due to the nonconvexity and discontinuity of the rank function.

The nuclear norm technique has gained success in many applications:

$$\text{rank}(X) \quad \Longrightarrow \quad \|X\|_* := \sum_{i=1}^{n} \sigma_i(X),$$

where $\sigma_1(X) \geq \cdots \geq \sigma_n(X)$ denote all the singular values of X.

Nuclear norm — convex envelope of the rank function over the unit ball of the spectral norm.



Nuclear norm estimator:

$$\widetilde{X} \in \arg\min_{\Omega \cap \mathcal{H}_+^n} \left\{ \|X\|_* \,:\, \|\mathcal{A}(X) - b\| \leq \delta \right\}$$

- Extra biased (for large singular values).
- It may fail when certain rows and / or columns are heavily sampled (Salakhutdinov and Srebro, 2010).
- It may not be rank consistent for general sampling even for the low-dimensional case (Bach, 2008).
- For correlation and density matrices, $\|X\|_* = \text{constant} \implies$ the nuclear norm technique's power is limited.

Outline

1. Problem formulation
2. Multi-stage convex relaxation approach
   - Exact penalty for the equivalent MPSDEC
   - A unified framework for multi-stage convex relaxation
3. The power of multi-stage convex relaxation approach
   - General error bound
   - Error bound comparison for the first two stages
   - Error bound comparison for the first k stages
4. Numerical results
5. Conclusion

Let $C(\mathbb{J})$ be the family of closed convex functions $\phi : \mathbb{J} \to \mathbb{R}$ with $\mathbb{J} \supseteq [0,1]$ and

- $\phi(0) = 0, \quad 0 < t^* := \arg\min_{0 \le t \le 1} \phi(t)$ and $|\phi'_-(0)| < +\infty$.

Rank minimization estimator is equivalent to the following mathematical program with the PSD equilibrium constraint (MPSDEC)

$$\begin{aligned} \min_{X \in \Omega, W \in \mathcal{H}^n} \quad & \langle I, \Phi(W) \rangle \\ \text{s.t.} \quad & \|\mathcal{A}(X) - b\| \le \delta \\ & \langle W, X \rangle = 0, \ X \succeq 0, \ 0 \preceq W \preceq I \end{aligned} \tag{1}$$

- Φ — the Löwner operator associated to $\phi \in C(\mathbb{J})$
- The bilinear constraint $\langle W, X \rangle = 0$ is the trouble maker.

Exact penalty for MPSDEC

To deal with $\langle W, X\rangle = 0$ efficiently, we consider the penalized version

$$\begin{aligned} \min_{X\in\Omega, W\in\mathcal{H}^n} \ & \langle I, \Phi(W)\rangle + \rho\langle W, X\rangle \\ \text{s.t.} \ & \|\mathcal{A}(X) - b\| \leq \delta \\ & X \succeq 0,\ 0 \preceq W \preceq I \end{aligned} \tag{2}$$

- The cost function of (2) has a desired structure though it is nonconvex
- The penalty problem (2) is exact in the following sense

Theorem

There exists $\bar{\rho} > 0$ such that when $\rho > \bar{\rho}$, the set of global optimal solutions of the penalty problem (2) coincides with that of the MPSDEC.

(S0) Select a function $\phi \in C(\mathbb{J})$. Let $W^0 := I$ and set $k := 1$. Choose $\rho_0 > 0$.

(S1) Solve the following weighted trace-norm minimization problem

$$X^k \in \underset{X \in \Omega \cap \mathcal{H}_+^n}{\arg\min} \left\{ \langle W^{k-1}, X \rangle : \; \|\mathcal{A}(X) - b\| \le \delta \right\}.$$

(S2) By the information of $\|X^k\|$ select a suitable $\rho_k = \delta_k \rho_{k-1}$ with $\delta_k \ge 1$.

(S3) Solve the following minimization problem

$$W^k \in \underset{0 \preceq W \preceq I}{\arg\min} \left\{ \langle I, \Phi(W) \rangle + \rho_k \langle W, X^k \rangle \right\}.$$

(S4) Set $k := k + 1$, and go to (S.1).

- The convex relaxation approach is solving (2) in an alternating way.
- The computational cost is the solution of the weighted trace-norm minimization problem since if $X^k = U^k \mathrm{Diag}(\lambda(X^k))(U^k)^T$ we have

$$W^k = U^k \mathrm{Diag}(w_1^k, \ldots, w_n^k)(U^k)^T.$$

1. $\phi_1(t) = -t \quad \Rightarrow w_i^k = \begin{cases} 0 & \text{if } \lambda_i(X^k) \geq \frac{1}{\rho_k} \\ 1 & \text{otherwise} \end{cases}$
2. $\phi_2(t) = \frac{a^2-1}{4} t^2 - at \quad \Rightarrow w_i^k = \min(\frac{2}{a^2-1} \max(a - \rho_k \lambda_i(X^k), 0), 1).$
3. $\phi_3(t) = t + \frac{1}{t+\epsilon} - \frac{1}{\epsilon} \quad \Rightarrow w_i^k = \max(\frac{1}{\sqrt{\rho_k \lambda_i(X^k)+1}} - \epsilon, 0).$
4. $\phi_4(t) = t - \ln(t+\epsilon) + \ln(\epsilon) \quad \Rightarrow w_i^k = \max(\frac{1}{\rho_k \lambda_i(X^k)+1} - \epsilon, 0).$

Outline

1. Problem formulation
2. Multi-stage convex relaxation approach
   - Exact penalty for the equivalent MPSDEC
   - A unified framework for multi-stage convex relaxation
3. The power of multi-stage convex relaxation approach
   - General error bound
   - Error bound comparison for the first two stages
   - Error bound comparison for the first k stages
4. Numerical results
5. Conclusion

Notation and assumptions

Recall that $\text{rank}(\overline{X}) = r$. Assume that $\overline{X} = [\overline{U}_1, \overline{U}_2]\text{Diag}(\lambda(\overline{X}))[\overline{U}_1, \overline{U}_2]^{\mathbb{T}}$, where $\overline{U}_1 \in \mathbb{O}^{n\times r}$ and $\overline{U}_2 \in \mathbb{O}^{n\times(n-r)}$. For any $X \in \mathcal{H}^n$, define

$$\mathcal{P}_T(X) = \overline{U}_1\overline{U}_1^{\mathbb{T}}X + X\overline{U}_1\overline{U}_1^{\mathbb{T}} - \overline{U}_1\overline{U}_1^{\mathbb{T}}X\overline{U}_1\overline{U}_1^{\mathbb{T}}.$$

Let $\mathcal{Q} = \mathcal{A}^*\mathcal{A}$. Define the restricted eigenvalues of $\mathcal{Q}$ by (Tong Zhang, 2010)

$$\vartheta_+(k) := \sup_{0<\text{rank}(X)\le k} \frac{\langle X, \mathcal{Q}(X)\rangle}{\|X\|_F^2}, \quad \vartheta_-(k) := \inf_{0<\text{rank}(X)\le k} \frac{\langle X, \mathcal{Q}(X)\rangle}{\|X\|_F^2}$$

Assumption: There exist a $c \in [0, 1)$ and an integer s ($1 \le s \le \frac{n-2r}{2}$) s.t.

$$\frac{\vartheta_+(s)}{\vartheta_-(2r+2s)} \le 1 + \frac{2c^2 s}{r}.$$

General error bound

Theorem

Let $\gamma_{k-1} := \frac{\|\mathcal{P}_T(W^{k-1})\|}{\lambda_{\min}((\overline{U}_2)^{\mathbb{T}} W^{k-1}\overline{U}_2)}$. *If* $0 \le \gamma_{k-1} \le \frac{1}{c}$, *then* $\|X^k - X^*\|_F \le \Xi(\gamma_{k-1})$ *with*

$$\Xi(t) := \sqrt{1 + \frac{rt^2}{2s}} \cdot \frac{2\sqrt{\vartheta_+(2r+s)}\,\delta}{(1-ct)\,\vartheta_-(2r+s)} \quad \text{for } \; t \ge 0.$$

- $\Xi(1)$ — the error bound for the nuclear norm estimator.
- If $c < 1 - \frac{1-\delta_{3r}(1+\frac{\sqrt{5}}{2})}{\sqrt{\frac{3}{2}}(1-\delta_{3r})}$, then we have the following inequality

$$\Xi(1) = \frac{\sqrt{6}\sqrt{\vartheta_+(3r)}\,\delta}{(1-c)\,\vartheta_-(3r)} \le \frac{\sqrt{6}\sqrt{1+\delta_{3r}}\,\delta}{(1-c)\,(1-\delta_{3r})} < \frac{3\delta\sqrt{1+\delta_{3r}}}{1-\delta_{3r}(1+\frac{\sqrt{5}}{2})}.$$

The last term is the bound given by Mohan and Fazel (2010).

Error bounds in the first two stages

Let ψ^* be the conjugate function of $\psi(t) := \begin{cases} \phi(t) & \text{if } t \in [0,1], \\ +\infty & \text{otherwise.} \end{cases}$ Define

$$a_0 := (\psi^*)'_+(0),\ a_1 := (\psi^*)'_-\big(-\rho_1\lambda_{r+1}(X^1)\big) \text{ and } b_1 := (\psi^*)'_+\big(-\rho_1\lambda_r(X^1)\big).$$

Theorem

Suppose that $\lambda_r(\overline{X}) = \beta\|X^1 - \overline{X}\|_F$ for some $\beta > 2$. If the parameter ρ_1 is chosen such that $\rho_1 \in \Big(0, \frac{-\phi'_+(0)}{\lambda_{r+1}(X^1)}\Big)$ and $\frac{b_1 + \frac{a_0}{\sqrt{2}}\ln\frac{\beta}{\beta-\sqrt{2}}}{a_1 - \frac{a_1}{\sqrt{2}}\ln\frac{\beta}{\beta-\sqrt{2}}} < 1$, then

$$\big\|X^2 - \overline{X}\big\|_F \le \Xi(\gamma_1) \le \Xi\left(\frac{b_1 + \frac{a_0}{\sqrt{2}}\ln\frac{\beta}{\beta-\sqrt{2}}}{a_1 - \frac{a_1}{\sqrt{2}}\ln\frac{\beta}{\beta-\sqrt{2}}}\right) < \Xi(1). \quad (3)$$

Table: The ratio of the error bounds given by different ϕ for the first two stages

	ρ_1	β	ratio	0	0.1	0.3	0.5	0.9
ϕ_1	$\left(\frac{1}{\lambda_r(X^1)}, \frac{1}{\lambda_{r+1}(X^1)}\right)$	3.0	$\frac{\Xi(\gamma_1)}{\Xi(1)}$	0.945	0.926	0.877	0.801	0.362
ϕ_2	$\left(\frac{2}{\lambda_r(X^1)}, \frac{1}{\lambda_{r+1}(X^1)}\right)$	4.0	$\frac{\Xi(\gamma_1)}{\Xi(1)}$	0.945	0.927	0.879	0.803	0.365
ϕ_3	$\left(\frac{3}{\lambda_r(X^1)}, \frac{0.8}{\lambda_{r+1}(X^1)}\right)$	5.0	$\frac{\Xi(\gamma_1)}{\Xi(1)}$	0.972	0.963	0.937	0.894	0.544
ϕ_4	$\left(\frac{3}{\lambda_r(X^1)}, \frac{0.8}{\lambda_{r+1}(X^1)}\right)$	5.0	$\frac{\Xi(\gamma_1)}{\Xi(1)}$	0.961	0.948	0.912	0.854	0.453
$\widehat{\phi}_4$	$\left(\frac{3}{\lambda_r(X^1)}, \frac{0.8}{\lambda_{r+1}(X^1)}\right)$	6.0	$\frac{\Xi(\gamma_1)}{\Xi(1)}$	0.993	0.991	0.984	0.972	0.829

Let $\widetilde{b}_k := (\psi^*)'_+\big[-\rho_k(\lambda_r(X^*) - \Xi(\widetilde{\gamma}_{k-1}))\big], \widetilde{a}_k := (\psi^*)'_-(-\rho_k\Xi(\widetilde{\gamma}_{k-1}))$,

$$\widetilde{\gamma}_k := \frac{\widetilde{b}_k + \frac{a_0}{\sqrt{2}} \ln \frac{\lambda_r(\overline{X})}{\lambda_r(\overline{X}) - \sqrt{2}\Xi(\widetilde{\gamma}_{k-1})}}{\widetilde{a}_k - \frac{\widetilde{a}_k}{\sqrt{2}} \ln \frac{\lambda_r(\overline{X})}{\lambda_r(\overline{X}) - \sqrt{2}\Xi(\widetilde{\gamma}_{k-1})}} \quad \text{for } k \geq 1 \text{ and } \widetilde{\gamma}_0 = 1.$$

Theorem

Suppose that $\lambda_r(\overline{X}) = \beta\Xi(1)$ *for some* $\beta > 2$. *If the parameter* ρ_1 *is chosen such that* $\rho_1 \in \left(0, \frac{-\phi'_+(0)}{\Xi(1)}\right)$ *and* $\frac{\widetilde{b}_1 + \frac{a_0}{\sqrt{2}} \ln \frac{\beta}{\beta - \sqrt{2}}}{\widetilde{a}_1 - \frac{\widetilde{a}_1}{\sqrt{2}} \ln \frac{\beta}{\beta - \sqrt{2}}} < 1$, *and the parameters* σ_k *are chosen such that* $\sigma_k \in \left[1, \frac{\Xi(\widetilde{\gamma}_{k-2})}{\Xi(\widetilde{\gamma}_{k-1})}\right]$, *then for all* $k \geq 2$ *we have* $\gamma_{k-1} \leq \widetilde{\gamma}_{k-1} < \widetilde{\gamma}_{k-2}$ *and*

$$\left\|X^k - \overline{X}\right\|_F \leq \Xi(\gamma_{k-1}) \leq \Xi(\widetilde{\gamma}_{k-1}) < \Xi(\widetilde{\gamma}_{k-2}). \qquad (4)$$

Table: The ratio of the error bounds obtained with ϕ_2 for the first five stages

	ρ_1	ratio	0	0.1	0.3	0.5	0.9
ϕ_2	$\left(\frac{0.7}{\Xi(1)}, \frac{1.2}{\Xi(1)}\right)$	$\Xi(\gamma_1)/\Xi(1)$	0.952	0.936	0.893	0.824	0.398
		$\Xi(\gamma_2)/\Xi(1)$	0.918	0.879	0.760	0.574	0.093
		$\Xi(\gamma_3)/\Xi(1)$	0.900	0.844	0.678	0.468	0.084
		$\Xi(\gamma_4)/\Xi(1)$	0.892	0.828	0.648	0.448	0.084
		$\Xi(\gamma_5)/\Xi(1)$	0.889	0.821	0.640	0.445	0.084

NUS National University of Singapore

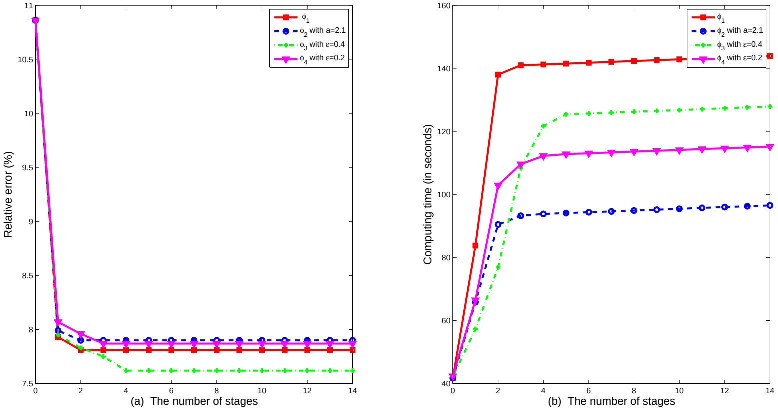


Figure: The performance of multi-stage convex relaxation with $p = 11940$

The performance with a low sample ratio

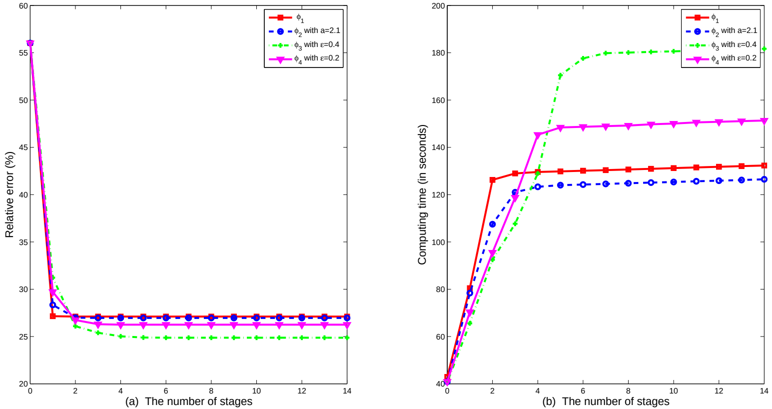


Figure: The performance of multi-stage convex relaxation with $p = 3980$

Low-rank correlation matrix recovery

Table: Performance for the correlation matrix recovery problems with $n = 1000$

r	off−diag	sr(%)	One-stage		Two-stage		Final result		
			relerr(rank)	time	relerr(rank)	time	k	relerr(rank)	time
5	0	1.60	8.19e-1(1000)	53.7	2.42e-1(19)	803	5	2.00e-1(5)	1656
	0	2.40	2.11e-1(998)	82.2	1.13e-1(5)	405	4	1.13e-1(5)	929
	0	3.99	9.97e-2(999)	107	7.09e-2(5)	669	4	7.06e-2(5)	1207
	100	1.60	7.94e-1(991)	58.0	2.26e-1(18)	788	5	1.88e-1(5)	1609
	100	2.40	2.08e-1(977)	92.8	1.11e-1(5)	415	4	1.11e-1(5)	846
	100	3.99	1.00e-1(703)	147	7.10e-2(6)	674	5	7.07e-2(5)	1309
10	0	3.59	6.54e-1(1000)	44.7	1.79e-1(21)	320	5	1.71e-1(10)	761
	0	5.38	1.56e-1(1000)	61.0	1.02e-1(10)	391	4	1.02e-1(10)	665
	0	8.96	8.98e-2(1000)	75.8	7.14e-2(10)	283	4	7.25e-2(10)	801
	100	3.59	6.29e-1(1000)	46.6	1.69e-1(17)	267	5	1.65e-1(10)	705
	100	5.38	1.48e-1(1000)	62.7	9.97e-2(10)	400	4	9.95e-2(10)	644
	100	8.97	8.91e-2(991)	82.3	7.09e-2(10)	290	4	7.22e-2(10)	787

Table: Performance for the low-rank density matrix recovery problems

r	n	sr(%)	One-stage		Two-stage		Final result		
			relerr(rank)	time	relerr(rank)	time	k	relerr(rank)	time
3	2^9	3.0	3.37e-1(31)	27.0	1.38e-1(3)	188	4	1.05e-1(3)	486
	2^9	3.0	2.48e-1(31)	26.5	1.31e-1(3)	138	4	1.04e-1(3)	317
	2^9	3.0	1.69e-1(31)	25.0	1.17e-1(3)	129	4	9.83e-2(3)	572
	2^{10}	2.0	1.31e-1(41)	281	9.31e-2(3)	780	4	7.26e-2(3)	1626
	2^{10}	2.00	1.53e-1(41)	243	1.07e-1(3)	935	4	8.38e-2(3)	1630
	2^{10}	2.5	1.05e-1(40)	340	7.97e-2(3)	761	4	6.24e-2(3)	1276
5	2^9	5.00	2.68e-1(42)	29.1	1.39e-1(5)	190	4	1.08e-1(5)	400
	2^9	4.00	4.81e-1(43)	22.6	1.65e-1(5)	449	5	1.37e-1(5)	1227
	2^9	3.00	5.56e-1(43)	17.4	2.92e-1(17)	220	5	2.21e-1(5)	1014

Table: Performance for the low-rank covariance matrix recovery with $n = 1000$

r	nfixed	sr(%)	One-stage		Two-stage		Final result		
			relerr(rank)	time	relerr(rank)	time	k	relerr(rank)	time
10	(200, 0)	3.58	4.56e-1(58)	239	2.11e-1(10)	358	4	2.11e-1(10)	578
	(0, 200)	3.58	4.57e-1(59)	188	2.15e-1(10)	301	4	2.14e-1(10)	630
	(200, 200)	3.58	4.23e-1(57)	218	2.16e-1(10)	349	4	2.16e-1(10)	580
	(200, 0)	7.16	1.37e-1(52)	170	9.83e-2(10)	317	4	9.69e-2(10)	444
	(0, 200)	7.16	1.41e-1(49)	66.3	1.00e-1(10)	211	4	9.84e-2(10)	329
	(200, 200)	7.16	1.37e-1(53)	164	9.72e-2(10)	304	4	9.62e-2(10)	463
20	(200, 0)	7.52	3.35e-1(86)	137	1.91e-1(20)	207	4	1.90e-1(20)	346
	(0, 200)	7.52	3.42e-1(88)	78.6	1.93e-1(20)	149	4	1.92e-1(20)	322
	(200, 200)	7.52	3.30e-1(87)	167	1.88e-1(20)	269	4	1.88e-1(20)	415
	(200, 0)	11.3	1.63e-1(79)	74.8	1.18e-1(20)	201	4	1.17e-1(20)	270
	(0, 200)	11.3	1.63e-1(78)	61.4	1.18e-1(20)	173	4	1.14e-1(20)	261
	(200, 200)	11.3	1.63e-1(80)	68.7	1.16e-1(20)	196	4	1.15e-1(20)	294

Outline

1. Problem formulation
2. Multi-stage convex relaxation approach
   - Exact penalty for the equivalent MPSDEC
   - A unified framework for multi-stage convex relaxation
3. The power of multi-stage convex relaxation approach
   - General error bound
   - Error bound comparison for the first two stages
   - Error bound comparison for the first k stages
4. Numerical results
5. Conclusion

Conclusion

- We proposed a multi-stage convex relaxation approach for the rank minimization problem by solving the exact penalty problem of its equivalent MPSDEC in an alternating way.
- Under a weaker condition than the RIP, we established a tighter error bound for the optimal solution of each subproblem to the global optimal solution of the original problem, and provided a quantitative analysis for the decrease of the error bound.
- The two-stage convex relaxation can improve the error bound given by the trace norm relaxation method at least 30% for those not nice problems.
- The general case can be done by using the theory on spectral operators of matrices.