Spectral Bundle Methods for Primal and Dual Semidefinite Programs

Feng-Yi Liao[1], Lijun Ding[2], Yang Zheng[1]

[1] Electrical and Computer Engineering, UC San Diego
[2] Industrial and Systems Engineering, Texax A & M University

IOS 2024, Houston, Texas
March 22, 2023

Semidefinite Program

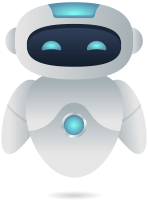
Primal SDP

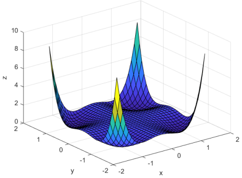
$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \ \ i = 1, \ldots, m, \\ & X \in \mathbb{S}^n_+. \end{aligned}$$

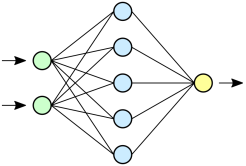
Dual SDP

$$\begin{aligned} \max_{y, Z} \quad & b^\mathsf{T} y \\ \text{subject to} \quad & Z + \sum_{i=1}^{m} A_i y_i = C, \\ & Z \in \mathbb{S}^n_+. \end{aligned}$$

- **Applications:** Control theory, combinatorial problem, polynomial optimization, neural network verification, robotics, etc.
- (Boyd et al. 1994, Sotirov 2012, Blekherman, Parrilo, and Thomas 2012, Lanckriet et al. 2014.)







Existing algorithms

- **Interior-point method:** Suitable for small or median-size problems
  - Solvers: MOSEK, SEDUMI, SDPA, SDP3
- **First-order method:** Speeds up the computation but suffers from moderate accuracy
  - Based on ADMM: SCS, CDCS, COSMO ('Donoghue et al. 2016, Zheng et al. 2020, Garstka et al. 2021)
- **Today's talk:** Spectral Bundle Method (SBM)

  Key references:

  1. Christoph Helmberg and Franz Rendl (2000). "A spectral bundle method for semidefinite programming". In: SIAM Journal on Optimization 10.3, pp. 673–696.
  2. Lijun Ding and Benjamin Grimmer (2023). "Revisiting Spectral Bundle Methods: Primal-Dual (Sub) linear Convergence Rates". In: SIAM Journal on Optimization 33.2, pp. 1305–1332.
  3. Feng-Yi Liao, Lijun Ding, and Yang Zheng (2023). "An Overview and Comparison of Spectral Bundle Methods for Primal and Dual Semidefinite Programs". In: arXiv preprint arXiv:2307.07651.

Spectral Bundle Method - Overview

- First Proposed by Helmberg and Rendl in 2000[1]
  - Well suited for combinatorial problems
  - Further developed by Helmberg et al. 2002, Helmberg et al. 2014
- Revisited by Lijun and Benjamin in 2023[2]
  - Discovered the linear convergence
  - SBM works well when the primal SDP has low-rank solutions
- However, all of the existing results focus on dual SDPs

$$\max_{y,Z} \; b^\mathsf{T} y$$
$$\text{subject to } \; Z + \sum_{i=1}^{m} A_i y_i = C,$$
$$Z \in \mathbb{S}^n_+.$$

- We show SBM also works for primal SDPs
- Primal SBM is very suitable for SDPs that have low-rank dual solutions

[1] Christoph Helmberg and Franz Rendl (2000). "A spectral bundle method for semidefinite programming". In: *SIAM Journal on Optimization* 10.3, pp. 673–696.

[2] Lijun Ding and Benjamin Grimmer (2023). "Revisiting Spectral Bundle Methods: Primal-Dual (Sub) linear Convergence Rates". In: *SIAM Journal on Optimization* 33.2, pp. 1305–1332.

Main results

- **Contribution 1**: We develop SBM for primal SDPs

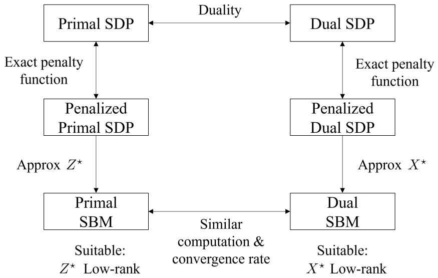
$$\begin{aligned} \min_X \quad & \langle C, X\rangle \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i,\ i = 1, \dots, m, \\ & X \in \mathbb{S}^n_+. \end{aligned} \iff \begin{aligned} \min_X \quad & \langle C, X\rangle + \rho \max\{\lambda_{\max}(-X), 0\} \\ \text{subject to} \quad & \langle A_i, X\rangle = b_i,\ i = 1, \dots, m. \end{aligned}$$

- **Contribution 2**: Extend the analysis of bundle methods from unconstrained to constrained problems

- **Contribution 3**: The algorithm enjoys linear convergence after some number of iterations

$$\mathrm{dist}(X_{t+1}, \mathcal{P}^\star) \leq \theta_t \cdot \mathrm{dist}(X_t, \mathcal{P}^\star), \text{ where } \theta_t < 1.$$

- **Contribution 4**: Comparison of primal and dual SBMs



Spectral Bundle Method

- **Penalized formulation**: under strong duality

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \; i = 1, \ldots, m, \\ & X \in \mathbb{S}^n_+. \end{aligned}$$

$$\begin{aligned} \min_{y,Z} \quad & -b^{\mathsf{T}} y \\ \text{subject to} \quad & Z + \sum_{i=1}^{m} A_i y_i = C, \\ & Z \in \mathbb{S}^n_+. \end{aligned}$$

$$\Big\Updownarrow \; \rho > \sup_{Z^\star} \operatorname{tr}(Z^\star)$$

$$\Big\Updownarrow \; \rho > \sup_{X^\star} \operatorname{tr}(X^\star)$$

$$\begin{aligned} \min_{X} \quad & \langle C, X \rangle + \rho \max\{\lambda_{\max}(-X), 0\} \\ \text{subject to} \quad & \langle A_i, X \rangle = b_i, \; i = 1, \ldots, m. \end{aligned}$$

$$\min_{y} \quad -b^{\mathsf{T}} y + \rho \max\left\{ \lambda_{\max}\left(\sum_{i=1}^{m} A_i y_i - C \right), 0 \right\}.$$

- Primal: constrained eigenvalue optimization problem
- Dual: unconstrained eigenvalue optimization problem
- We are going to iteratively approximate the nonsmooth nonlinear penalty function

$$\max\{\lambda_{\max}(-X), 0\}$$

Spectral Bundle Method

- Consider

$$\min_{X \in \mathcal{X}_0} \; f(X) = \langle C, X \rangle + \rho \max\{\lambda_{\max}(-X), 0\}$$

Proximal mapping

$$\boxed{\min_{X \in \mathcal{X}_0} \; \hat{f}_t(X) + \frac{\alpha}{2}\|X - X_t\|^2}$$

Update

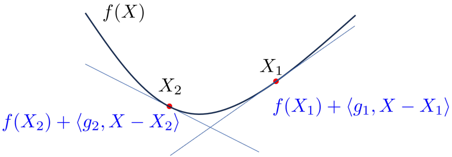
$$\boxed{\hat{f}_t \to \hat{f}_{t+1} \text{ and } X_t \to X_{t+1}}$$

- Lower model:

$$\hat{f}_t(X) \le f(X), \;\; \forall X \in \mathcal{X}_0$$

- Piece-wise linear model: given X_i (one eigenvector (subgradient) at a time)

$$f(X_i) + \langle g_i, X - X_i \rangle = \left\langle v_i v_i^{\mathsf{T}}, -X \right\rangle \le \max\{\lambda_{\max}(-X), 0\}$$



- In general, the iteration complexity[3] is

Lipschitz	Lipschitz + Quadratic Growth
$\mathcal{O}(\frac{1}{\epsilon^3})$ or $\mathcal{O}(\frac{1}{\epsilon^2})$	$\mathcal{O}(\frac{1}{\epsilon})$

- **Can we do better? What if we compute $r \geq 1$ eigenvectors at a time?**
- Lower approximation model using r eigenvectors:

$$\left\langle v_i v_i^{\mathsf{T}}, -X \right\rangle \leq \max_{S \in \mathbb{S}^r_+, \mathrm{tr}(S) \leq 1} \left\langle PSP^{\mathsf{T}}, -X \right\rangle \leq \max\{\lambda_{\max}(-X), 0\},$$

  where $P \in \mathbb{R}^{n \times r}$ contains r top eigenvectors of $-X_i$.
- This contains **infinitely** many lower linear approximation at a time
- **Does it improve the iteration complexity?**
  - **Yes!** Choosing a certain number of r improves the iteration complexity.
  - $\mathcal{O}(1/\epsilon) \rightarrow \mathcal{O}(\log(1/\epsilon))$

[3] Mateo Díaz and Benjamin Grimmer (2023). "Optimal convergence rates for the proximal bundle method". In: *SIAM Journal on Optimization* 33.2, pp. 424–454.

Spectral Bundle Method

Theorem (Simplified, Sublinear convergence) Primal SBM with $r \geq 1$ eigenvector produces iterates X_t, Z_t, and y_t that satisfy

$$\begin{aligned} \text{cost value gap:} \quad & f(X_t) - f^\star \leq \epsilon, \\ \text{approximate primal feasibility:} \quad & \mathcal{A}(X_t) - b = 0, \quad \lambda_{\min}(X_t) \geq -\epsilon, \\ \text{approximate dual feasibility:} \quad & \|Z_t + \mathcal{A}^*(y_t) - C\|^2 \leq \epsilon, \quad Z_t \succeq 0, \\ \text{approximate primal-dual optimality:} \quad & |\langle C, X_t\rangle - \langle b, y_t\rangle| \leq \sqrt{\epsilon}, \end{aligned}$$

in at most $\mathcal{O}(1/\epsilon^3)$ iterations.

- If additionally, strict complementarity holds ($\mathrm{rank}(X^\star) + \mathrm{rank}(Z^\star) = n$), then these conditions hold in at most $\mathcal{O}(1/\epsilon)$ iterations.

- Assuming
  1. strict complementarity ($\mathrm{rank}(X^\star) + \mathrm{rank}(Z^\star) = n$)
  2. bounded solution set

  then given any sublevel set $[f \leq f^\star + \epsilon]$, there exists $\mu > 0$ such that

$$\mu \cdot \mathrm{dist}^2(X, \mathcal{P}^\star) \leq f(X) - f(X^\star), \ \forall X \in [f \leq f^\star + \epsilon].$$

Theorem. (Simplified, Linear convergence) Under certain regularity conditions, primal SBM with $r \geq \mathrm{rank}(Z^\star)$ eigenvectors generates $\{X_t\}$ that satisfies

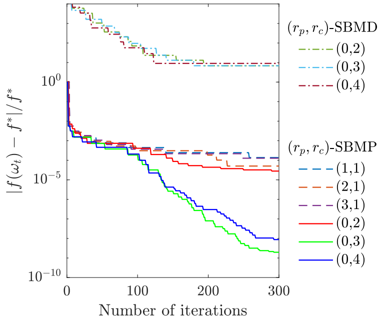
$$\mathrm{dist}(X_{t+1}, \mathcal{P}^\star) \leq \theta_t \cdot \mathrm{dist}(X_t, \mathcal{P}^\star), \text{ with } \theta_t < 1,$$

for every $t \geq T_0$.

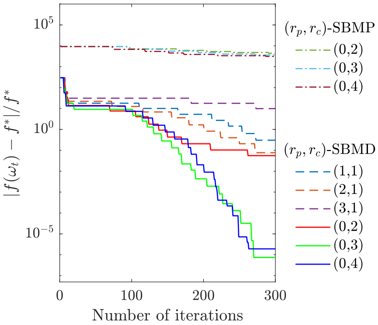
- If $\mathrm{rank}(Z^\star)$ is small, r can be chosen small! $\Rightarrow$ Faster computation!
- Similar results also hold for the dual SBM (Lijun and Benjamin 2023)
- Key difference: $r \geq \mathrm{rank}(X^\star)$

Descriptions	Primal	Dual
Primal feasibility	$\mathcal{A}(X_t) = b, \lambda_{\min}(X_t) \geq -\epsilon$	$\lVert\mathcal{A}(X_t) - b\rVert^2 \leq \epsilon, X_t \succeq 0$
Dual feasibility	$\lVert Z_t + \mathcal{A}^*(y_t) - C\rVert^2 \leq \epsilon, Z_t \succeq 0$	$C - \mathcal{A}^*(y_t) = Z_t, \lambda_{\min}(Z_t) \geq -\epsilon$
Duality gap	$\lvert\langle C, X_t\rangle - \langle b, y_t\rangle\rvert \leq \sqrt{\epsilon}$	$\lvert\langle C, X_t\rangle - \langle b, y_t\rangle\rvert \leq \sqrt{\epsilon}$
Well-suited	$Z^\star$ low rank	$X^\star$ low rank

Open-source implementation: `https://github.com/soc-ucsd/SpecBM`



(a) An SDP instance with $\mathrm{rank}(Z^\star) = 3$



(b) An SDP instance with $\mathrm{rank}(X^\star) = 3$

Figure: Random generated SDPs with dimension $n = 1000$.

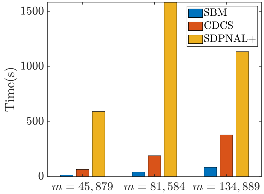
- SBMP (primal) performs well when $Z^\star$ is low rank
- SBMD (dual) performs well when $X^\star$ is low rank
- A significant improvement when $r \geq 3 = \mathrm{rank}(Z^\star) = \mathrm{rank}(X^\star)$

Numerical Experiments

$$\min_{x \in \mathbb{R}^d} \quad q(x)$$

$$\text{subject to} \quad x^{\mathsf{T}} x = 1$$

- $q(x)$ is a multi-variables polynomial
- Sum-of-square relaxation to a standard primal SDP
- The dual problem has low-rank solutions empirically
- Accuracy is set $\epsilon < 10^{-4}$



Dimension	Case	Metric	SDPT3	MOSEK	CDCS	SDPNAL+	SBMP
d: 30, n: 496, m: 45,879	q_1	ϵ_p	oom	9.2e−11	9.9e−5	2.1e−11	5.9e−14
		ϵ_d		2.0e−9	4.1e−5	9.5e−5	9.2e−5
		ϵ_g		5.9e−12	9.0e−6	1.0e−2	6.8e−6
		cost		−25.074	−25.073	−42.850	−25.074
		time		887.1	100.9	1209	**17.5**
	q_2	ϵ_p	oom	8.2e−10	1.5e−5	4.7e−5	5.4e−14
		ϵ_d		1.9e−8	9.3e−5	8.1e−5	8.9e−5
		ϵ_g		2.9e−10	3.9e−5	6.7e−3	5.3e−5
		cost		−32.136	−32.137	−32.137	−32.135
		time		583	45.1	17.3	**12.8**
	q_3	ϵ_p	oom	9.1e − 9	1.0e−4	3.2e−16	1.9e−14
		ϵ_d		2.0e−7	1.5e−6	5.2e−4	8.7e−5
		ϵ_g		4.9e−9	2.4e−6	3.5e−3	2.2e−5
		cost		2.023	2.023	2.021	2.023
		time		645.2	55.3	548.6	**16.9**
d: 35, n: 666, m: 81,584	q_1	ϵ_p	oom	oom	9.6e−5	1.1e−15	9.9e−14
		ϵ_d			1.9e−5	6.8e−5	9.9e−5
		ϵ_g			2.1e−5	2.7e−4	9.4e−6
		cost			−30.053	−30.070	−30.050
		time			289.6	3449.4	**41.4**
	q_2	ϵ_p	oom	oom	2.2e−5	4.2e−5	1.3e−13
		ϵ_d			9.4e−5	9.5e−5	8.8e−5
		ϵ_g			7.7e−5	2.4e−3	2.0e−5
		cost			−37.100	−37.221	−37.091
		time			112.3	100.5	**42.7**
	q_3	ϵ_p	oom	oom	1.0e−4	9.3e−13	1.9e−14
		ϵ_d			6.9e−5	6.5e−7	9.4e−5
		ϵ_g			3.5e−5	6.2e−3	2.5e−6
		cost			2.121	2.039	2.121
		time			170.1	1208.4	**44.5**
d: 40, n: 861, m: 134,889	q_1	ϵ_p	oom	oom	1.0e−4	3.6e−12	7.4e−14
		ϵ_d			3.2e−6	1.0e−4	9.6e−5
		ϵ_g			1.4e−5	4.9e−2	5.2e−6
		cost			−35.037	−47.792	−35.034
		time			738.9	1881.1	**84.8**
	q_2	ϵ_p	oom	oom	1.7e−5	3.3e−14	1.3e−13
		ϵ_d			9.3e−5	9.9e−5	5.5e−5
		ϵ_g			5.8e−5	2.8e−5	1.6e−5
		cost			−42.041	−41.997	−42.050
		time			254.9	187.6	**82.1**
	q_3	ϵ_p	oom	oom	1.0e−4	9.1e−16	2.0e−14
		ϵ_d			1.6e−6	7.5e−5	8.2e−5
		ϵ_g			4.9e−5	2.4e−3	1.0e−4
		cost			2.681	2.695	2.681
		time			144.5	1339.6	**93.1**

Takeaways

- Penalized formulation:

$$f(X) = \langle C, X \rangle + \rho \max\{\lambda_{\max}(-X), 0\} \qquad (\rho > \max_{Z^\star} \operatorname{tr}(Z^\star))$$

$$f(y) = -b^{\mathsf{T}} y + \rho \max\left\{\lambda_{\max}\left(\sum_{i=1}^{m} A_i y_i - C\right), 0\right\} \qquad (\rho > \max_{X^\star} \operatorname{tr}(X^\star))$$

- Sublinear $\longrightarrow$ Linear!

$$\operatorname{dist}(X_{t+1}, \mathcal{P}_\star) \le \theta_t \cdot \operatorname{dist}(X_t, \mathcal{P}_\star), \text{ with } \theta_t < 1.$$

- Comparison

SBM	Primal	Dual
Usage	$Z^\star$ low rank	$X^\star$ low rank

- Choose the right formulation!
- Check out our paper: Feng-Yi Liao, Lijun Ding, and Yang Zheng (2023). "An Overview and Comparison of Spectral Bundle Methods for Primal and Dual Semidefinite Programs". In: arXiv preprint arXiv:2307.07651.

Thank you for your attention!

Q & A