An accelerated first-order method for a class of semidefinite programs

Alex L. Wang,Centrum Wiskunde & Informatica
Fatma Kılınç-Karzan,Carnegie Mellon University

Workshop on Semidefinite and Polynomial Optimization
August 31, 2022

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

  - Nonconvex, generally NP-hard

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

  - Nonconvex, generally NP-hard, but much smaller nk

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

  - Nonconvex, generally NP-hard, but much smaller nk
- **Sneak peek**

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0,\ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

  - Nonconvex, generally NP-hard, but much smaller nk
- **Sneak peek**
  - SDPs with “low rank structure” can be reformulated as “easy” QMPs

---

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$

  - Nonconvex, generally NP-hard, but much smaller nk
- **Sneak peek**
  - SDPs with "low rank structure" can be reformulated as "easy" QMPs
  - First-order method (FOM) for "easy" QMPs

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

5-minute flash talk

- Semidefinite program

$$(\mathbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

  - Convex, useful from modeling perspective, but problem is very large $\binom{n+k+1}{2}$
- Quadratic matrix program

$$(\mathbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

where q_i are **quadratic matrix functions**: $q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$
  - Nonconvex, generally NP-hard, but much smaller nk
- **Sneak peek**
  - SDPs with "low rank structure" can be reformulated as "easy" QMPs
  - First-order method (FOM) for "easy" QMPs
  - $\longrightarrow$ Storage-optimal, low complexity FOM for SDPs with "low rank structure"

Related: Beck [2007], Beck et al. [2012], Burer and Monteiro [2003], Ding et al. [2021], Yurtsever et al. [2021]

Outline for today's talk

- **Quadratic matrix programs**

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs
  - The Burer–Monteiro method + drawbacks of symmetry

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs
  - The Burer–Monteiro method + drawbacks of symmetry
- **FOM for easy QMPs**

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs
  - The Burer–Monteiro method + drawbacks of symmetry
- **FOM for easy QMPs**
  - Algorithmic ideas

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs
  - The Burer–Monteiro method + drawbacks of symmetry
- **FOM for easy QMPs**
  - Algorithmic ideas
  - Computational results

Outline for today's talk

- **Quadratic matrix programs**
  - "Easy" QMPs: strict complementarity
  - The SDP relaxation of a QMP
- **Relating SDPs to QMPs**
  - "Low-rank structure": k-exact SDPs
  - The Burer–Monteiro method + drawbacks of symmetry
- **FOM for easy QMPs**
  - Algorithmic ideas
  - Computational results
- **Conclusion**

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n\times k}} \{q_0(X) : \; q_i(X) = 0, \, \forall i \in [m]\}$$

$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X\in\mathbb{R}^{n\times k}} \{q_0(X) : \; q_i(X) = 0, \; \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i$$

- If $k = 1$, $$\min_{x\in\mathbb{R}^n} \{q_0(x) : \; q_i(x) = 0, \; \forall i \in [m]\}$$

  where q_i are quadratic functions

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) :\ q_i(X) = 0,\ \forall i \in [m]\}$$

$$q_i(X) = \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i$$

- If $k = 1$, $$\min_{x \in \mathbb{R}^n} \{q_0(x) :\ q_i(x) = 0,\ \forall i \in [m]\}$$ where q_i are quadratic functions
- QMPs are hard

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) : \ q_i(X) = 0, \ \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

- If $k = 1$, $$\min_{x \in \mathbb{R}^n} \{q_0(x) : \ q_i(x) = 0, \ \forall i \in [m]\}$$

  where q_i are quadratic functions
- QMPs are hard
  - Binary program $\longrightarrow x_1(1 - x_1) = 0$

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) : \; q_i(X) = 0, \; \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

- If $k = 1$,
  $$\min_{x \in \mathbb{R}^n} \{q_0(x) : \; q_i(x) = 0, \; \forall i \in [m]\}$$
  where q_i are quadratic functions
- QMPs are hard
  - Binary program $\longrightarrow x_1(1 - x_1) = 0$
  - Polynomial optimization problem $\longrightarrow x_1 x_2 = z_{1,2}$

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) : \, q_i(X) = 0, \, \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

- If $k = 1$, $$\min_{x \in \mathbb{R}^n} \{q_0(x) : \, q_i(x) = 0, \, \forall i \in [m]\}$$
  where q_i are quadratic functions
- QMPs are hard
  - Binary program $\longrightarrow x_1(1 - x_1) = 0$
  - Polynomial optimization problem $\longrightarrow x_1 x_2 = z_{1,2}$
- The SDP relaxation of a QMP?

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) : q_i(X) = 0, \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

- If $k = 1$, $$\min_{x \in \mathbb{R}^n} \{q_0(x) : q_i(x) = 0, \forall i \in [m]\}$$
  where q_i are quadratic functions
- QMPs are hard
  - Binary program $\longrightarrow x_1(1 - x_1) = 0$
  - Polynomial optimization problem $\longrightarrow x_1 x_2 = z_{1,2}$
- The SDP relaxation of a QMP? Next slide.

Related: Beck [2007], Beck et al. [2012]

Quadratic matrix programs

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n \times k}} \{q_0(X) : q_i(X) = 0, \forall i \in [m]\}$$
$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

- If $k = 1$,
$$\min_{x \in \mathbb{R}^n} \{q_0(x) : q_i(x) = 0, \forall i \in [m]\}$$
where q_i are quadratic functions
- QMPs are hard
  - Binary program $\longrightarrow x_1(1 - x_1) = 0$
  - Polynomial optimization problem $\longrightarrow x_1 x_2 = z_{1,2}$
- The SDP relaxation of a QMP? Next slide.
- QMP is “easy” if the SDP relaxation solves it

Related: Beck [2007], Beck et al. [2012]

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$q_i(X) = \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$\begin{aligned} q_i(X) &= \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i \\ &= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\intercal & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\intercal & X \\ X^\intercal & I_k \end{pmatrix} \right\rangle \end{aligned}$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$\begin{aligned} q_i(X) &= \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i \\ &= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\top & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\top & X \\ X^\top & I_k \end{pmatrix} \right\rangle =: \langle M_i, Y(X)\rangle \end{aligned}$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$\begin{aligned} q_i(X) &= \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i \\ &= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\intercal & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\intercal & X \\ X^\intercal & I_k \end{pmatrix} \right\rangle =: \langle M_i, Y(X)\rangle \end{aligned}$$

$$(\textbf{QMP}) = \min_{X \in \mathbb{R}^{n\times k}} \{q_0(X) : \, q_i(X) = 0, \, \forall i \in [m]\}$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$
\begin{aligned}
q_i(X) &= \langle X, A_i X \rangle + 2 \langle B_i, X \rangle + c_i \\
&= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\intercal & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\intercal & X \\ X^\intercal & I_k \end{pmatrix} \right\rangle =: \langle M_i, Y(X) \rangle
\end{aligned}
$$

$$
\begin{aligned}
(\textbf{QMP}) &= \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : \; q_i(X) = 0, \; \forall i \in [m] \} \\
&= \min_{X \in \mathbb{R}^{n \times k}} \{ \langle M_0, Y(X) \rangle : \; \langle M_i, Y(X) \rangle = 0, \; \forall i \in [m] \}
\end{aligned}
$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

$$\begin{aligned} q_i(X) &= \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i \\ &= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\intercal & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\intercal & X \\ X^\intercal & I_k \end{pmatrix} \right\rangle =: \langle M_i, Y(X)\rangle \end{aligned}$$

$$\begin{aligned} (\textbf{QMP}) &= \min_{X\in\mathbb{R}^{n\times k}} \{q_0(X) : q_i(X) = 0,\ \forall i \in [m]\} \\ &= \min_{X\in\mathbb{R}^{n\times k}} \{\langle M_0, Y(X)\rangle : \langle M_i, Y(X)\rangle = 0,\ \forall i \in [m]\} \\ &\geq \min_{Y\in\mathbb{S}^{n+k}} \left\{ \langle M_0, Y\rangle : \begin{array}{l} \langle M_i, Y\rangle = 0,\ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} =: (\textbf{SDP}) \end{aligned}$$

Two views of the SDP relaxation of a QMP

- Lift and relax:

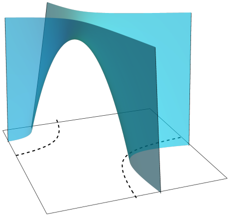
$$\begin{aligned} q_i(X) &= \langle X, A_i X\rangle + 2\langle B_i, X\rangle + c_i \\ &= \left\langle \begin{pmatrix} A_i & B_i \\ B_i^\intercal & \frac{c_i}{k} I_k \end{pmatrix}, \begin{pmatrix} XX^\intercal & X \\ X^\intercal & I_k \end{pmatrix} \right\rangle =: \langle M_i, Y(X)\rangle \end{aligned}$$

$$\begin{aligned} (\textbf{QMP}) &= \min_{X\in\mathbb{R}^{n\times k}} \{q_0(X) : q_i(X) = 0,\ \forall i \in [m]\} \\ &= \min_{X\in\mathbb{R}^{n\times k}} \{\langle M_0, Y(X)\rangle : \langle M_i, Y(X)\rangle = 0,\ \forall i\in[m]\} \\ &\geq \min_{Y\in\mathbb{S}^{n+k}} \left\{ \langle M_0, Y\rangle : \begin{array}{c} \langle M_i, Y\rangle = 0,\ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} =: (\textbf{SDP}) \end{aligned}$$

- **Standing assumption**: (**QMP**) feasible and (**SDP**) dual strictly feasible

Two views of the SDP relaxation of a QMP cont.

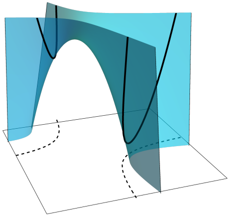
- Lagrangian aggregation:



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

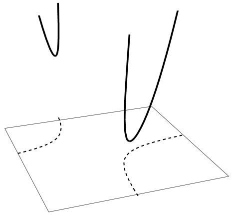
- Lagrangian aggregation:



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation:

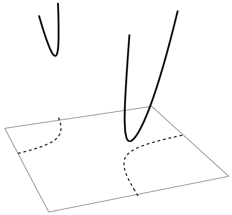


$$\min_{X\in\mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

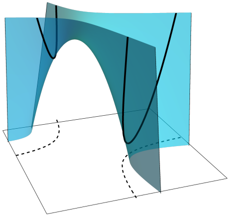
$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$

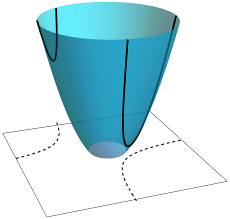


$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$



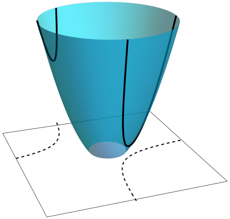
$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^m \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$

$$q(\gamma, X) = \langle X, A(\gamma)X \rangle + 2\langle B(\gamma), X \rangle + c(\gamma)$$



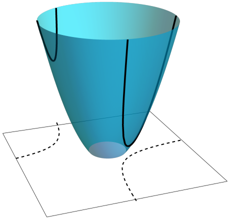
$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$

$$q(\gamma, X) = \langle X, A(\gamma)X \rangle + 2 \langle B(\gamma), X \rangle + c(\gamma)$$

- $$(\textbf{QMP}) \geq \sup_{\gamma \in \mathbb{R}^m} \inf_{X \in \mathbb{R}^{n \times k}} q(\gamma, X)$$



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

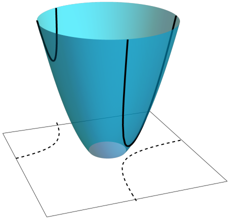
Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$

$$q(\gamma, X) = \langle X, A(\gamma)X \rangle + 2\langle B(\gamma), X \rangle + c(\gamma)$$

- $$(\textbf{QMP}) \geq \sup_{\gamma \in \mathbb{R}^m} \inf_{X \in \mathbb{R}^{n \times k}} q(\gamma, X)$$
$$= \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X) = (\textbf{Lagr.})$$



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

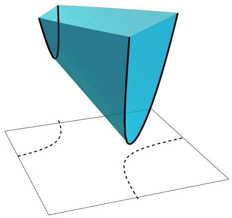
Two views of the SDP relaxation of a QMP cont.

- Lagrangian aggregation: $q_i(\bar{X}) = 0$ and $\gamma \in \mathbb{R}^m$

$$q_0(\bar{X}) = q_0(\bar{X}) + \sum_{i=1}^{m} \gamma_i q_i(\bar{X}) =: q(\gamma, \bar{X})$$

$$q(\gamma, X) = \langle X, A(\gamma) X \rangle + 2 \langle B(\gamma), X \rangle + c(\gamma)$$

- $$(\textbf{QMP}) \geq \sup_{\gamma \in \mathbb{R}^m} \inf_{X \in \mathbb{R}^{n \times k}} q(\gamma, X)$$
$$= \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X) = (\textbf{Lagr.})$$



$$\min_{X \in \mathbb{R}^2} \{q_0(X) : q_1(X) = 0\}$$

Two views of the SDP relaxation of a QMP cont.

- Suppose (**SDP**) feasible and (**SDP**) dual strictly feasible. Then,

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y = \begin{pmatrix} * & X \\ X^\intercal & I_k \end{pmatrix} \succeq 0 \end{array} \right\}$$

$$\begin{array}{ccc} \| & & \| \end{array}$$

$$(\textbf{Lagr.}) = \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma :\, A(\gamma) \succeq 0} q(\gamma, X)$$

Easy instances of QMPs

$$(\textbf{QMP}) \geq \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle = 0, \ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} = \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma :\, A(\gamma) \succeq 0} q(\gamma, X)$$

Easy instances of QMPs

$$(\textbf{QMP}) \geq \min_{Y\in\mathbb{S}^{n+k}} \left\{ \langle M_0, Y\rangle : \begin{array}{l} \langle M_i, Y\rangle = 0,\ \forall i\in[m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} = \min_{X\in\mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

Definition

QMP satisfies strict complementarity if, equivalently,

Easy instances of QMPs

$$(\textbf{QMP}) \geq \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle = 0,\ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} = \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X)$$

Definition

QMP satisfies strict complementarity if, equivalently,

- dual to (**Lagr.**) has optimal solution γ^* satisfying $A(\gamma^*) \succ 0$

Easy instances of QMPs

$$(\textbf{QMP}) \geq \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y\rangle : \begin{array}{l} \langle M_i, Y\rangle = 0,\ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} = \min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

Definition

QMP satisfies strict complementarity if, equivalently,

- dual to (**Lagr.**) has optimal solution γ^* satisfying $A(\gamma^*) \succ 0$
- dual to (**SDP**) is strictly feasible and has optimal solution with rank n

Easy instances of QMPs

$$(\textbf{QMP}) \geq \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{c} \langle M_i, Y \rangle = 0, \ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \end{array} \right\} = \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X)$$

Definition

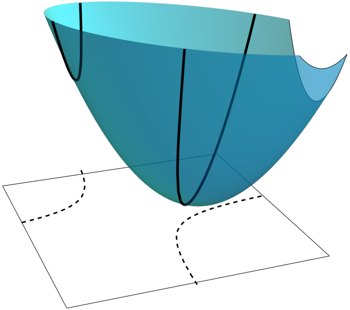
QMP satisfies strict complementarity if, equivalently,

- dual to (**Lagr.**) has optimal solution γ^* satisfying $A(\gamma^*) \succ 0$
- dual to (**SDP**) is strictly feasible and has optimal solution with rank n

- Then, QMP has unique optimal solution X^* and X^* is unique solution (**Lagr.**)

Easy instances of QMPs cont.

X^* is unique solution (**Lagr.**) $\quad A(\gamma^*) \succ 0$



$$X^* = \arg\min_{X \in \mathbb{R}^{n \times k}} q(\gamma^*, X)$$

Structural assumptions on SDP

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$
$$= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}$$

Related: Alizadeh et al. [1997], Ding et al. [2021]

Structural assumptions on SDP

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$
$$= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}$$

- What SDPs can be rewritten as QMPs with strict complementarity?

Related: Alizadeh et al. [1997], Ding et al. [2021]

Structural assumptions on SDP

$$
\begin{aligned}
(\textbf{SDP}) &= \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\} \\
&= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}
\end{aligned}
$$

- What SDPs can be rewritten as QMPs with strict complementarity?

Definition

An SDP is k-exact QMP-like if

Related: Alizadeh et al. [1997], Ding et al. [2021]

Structural assumptions on SDP

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$
$$= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}$$

- What SDPs can be rewritten as QMPs with strict complementarity?

Definition

An SDP is k-exact QMP-like if

- Strong duality holds, both are solvable, there exists Y^* and γ^*

Related: Alizadeh et al. [1997], Ding et al. [2021]

Structural assumptions on SDP

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$
$$= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}$$

- What SDPs can be rewritten as QMPs with strict complementarity?

Definition

An SDP is k-exact QMP-like if

- Strong duality holds, both are solvable, there exists Y^* and γ^*
- Strict complementarity: $\mathrm{rank}(Y^*) = k$ and $\mathrm{rank}(M(\gamma^*)) = n$

Related: Alizadeh et al. [1997], Ding et al. [2021]

Structural assumptions on SDP

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$
$$= \sup_{\gamma \in \mathbb{R}^m} \{ d^\intercal \gamma : M(\gamma) \succeq 0 \}$$

- What SDPs can be rewritten as QMPs with strict complementarity?

Definition

An SDP is k-exact QMP-like if

- Strong duality holds, both are solvable, there exists Y^* and γ^*
- Strict complementarity: $\operatorname{rank}(Y^*) = k$ and $\operatorname{rank}(M(\gamma^*)) = n$
- Know $Y^*_W \succ 0$ for some k-dimensional subspace W

Related: Alizadeh et al. [1997], Ding et al. [2021]

From k-exact SDP to easy QMP

- WLOG $W =$ last k-coordinates and $Y_W^* = I_k$, then

From k-exact SDP to easy QMP

- WLOG $W =$ last k-coordinates and $Y^*_W = I_k$, then

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{c} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \\ \operatorname{rank}(Y) = k \end{array} \right\}$$

From k-exact SDP to easy QMP

- WLOG W = last k-coordinates and $Y^*_W = I_k$, then

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{c} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \\ \operatorname{rank}(Y) = k \end{array} \right\}$$

$$= \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : \ q_i(X) = 0, \ \forall i \in [m] \}$$

From k-exact SDP to easy QMP

- WLOG $W =$ last k-coordinates and $Y^*_W = I_k$, then

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{c} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y = \begin{pmatrix} * & * \\ * & I_k \end{pmatrix} \succeq 0 \\ \operatorname{rank}(Y) = k \end{array} \right\}$$

$$= \min_{X \in \mathbb{R}^{n \times k}} \{ q_0(X) : q_i(X) = 0, \ \forall i \in [m] \}$$

$$= \min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma :\, A(\gamma) \succeq 0} q(\gamma, X)$$

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

- Suppose SDP is k-exact

$$(\textbf{SDP}) = \min_{\tilde{X} \in \mathbb{R}^{(n+k) \times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0, \ \forall i \ \right\}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

- Suppose SDP is k-exact

$$(\textbf{SDP}) = \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0, \ \forall i \right\}$$

$$= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0, \ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$

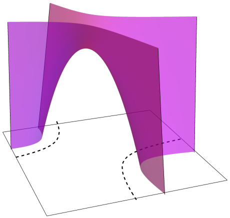
- Suppose SDP is k-exact

$$\begin{aligned}
(\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k) \times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0, \ \forall i \ \right\} \\
&= \min_{\tilde{X} \in \mathbb{R}^{(n+k) \times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \\
&= \min_{\tilde{X} \in \mathbb{R}^{(n+k) \times k}} \sup_{\gamma \in \mathbb{R}^m : \, M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma
\end{aligned}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$



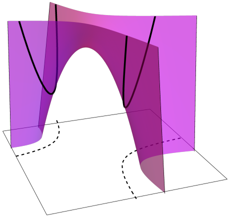
- Suppose SDP is k-exact

$$\begin{aligned}(\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0,\ \forall i \right\} \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma)\tilde{X} \right\rangle + d^\intercal \gamma \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m :\, M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma)\tilde{X} \right\rangle + d^\intercal \gamma\end{aligned}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$



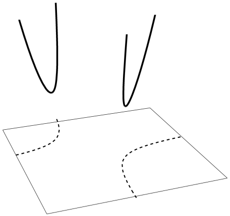
- Suppose SDP is k-exact

$$\begin{aligned}(\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0,\ \forall i \right\} \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m :\, M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma\end{aligned}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y\in\mathbb{S}^{n+k}} \left\{ \langle M_0, Y\rangle : \begin{array}{l} \langle M_i, Y\rangle + d_i = 0,\ \forall i\in[m] \\ Y\succeq 0 \end{array} \right\}$$

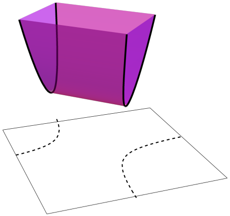


- Suppose SDP is k-exact

$$\begin{aligned}(\textbf{SDP}) &= \min_{\tilde{X}\in\mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0\tilde{X}\right\rangle : \ \left\langle \tilde{X}, M_i\tilde{X}\right\rangle + d_i = 0,\ \forall i \right\} \\ &= \min_{\tilde{X}\in\mathbb{R}^{(n+k)\times k}} \sup_{\gamma\in\mathbb{R}^m} \left\langle \tilde{X}, M(\gamma)\tilde{X}\right\rangle + d^\intercal\gamma \\ &= \min_{\tilde{X}\in\mathbb{R}^{(n+k)\times k}} \sup_{\gamma\in\mathbb{R}^m:\, M(\gamma)\succeq 0} \left\langle \tilde{X}, M(\gamma)\tilde{X}\right\rangle + d^\intercal\gamma\end{aligned}$$

Related: Burer and Monteiro [2003]

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$



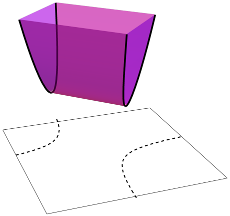
- Suppose SDP is k-exact

$$\begin{aligned} (\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0,\ \forall i \right\} \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m :\, M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \end{aligned}$$

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$



- Suppose SDP is k-exact

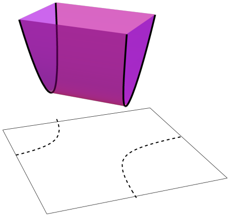
$$\begin{aligned}(\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \ \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0,\ \forall i \right\} \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \\ &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m :\ M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma\end{aligned}$$

- Too much symmetry!

Related: Burer and Monteiro [2003]

The Burer–Monteiro method

$$(\textbf{SDP}) = \min_{Y \in \mathbb{S}^{n+k}} \left\{ \langle M_0, Y \rangle : \begin{array}{l} \langle M_i, Y \rangle + d_i = 0,\ \forall i \in [m] \\ Y \succeq 0 \end{array} \right\}$$



- Suppose SDP is k-exact

$$\begin{aligned}
(\textbf{SDP}) &= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \left\{ \left\langle \tilde{X}, M_0 \tilde{X} \right\rangle : \left\langle \tilde{X}, M_i \tilde{X} \right\rangle + d_i = 0,\ \forall i \right\} \\
&= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma \\
&= \min_{\tilde{X} \in \mathbb{R}^{(n+k)\times k}} \sup_{\gamma \in \mathbb{R}^m :\, M(\gamma) \succeq 0} \left\langle \tilde{X}, M(\gamma) \tilde{X} \right\rangle + d^\intercal \gamma
\end{aligned}$$

- Too much symmetry! We set $\tilde{X} = \begin{pmatrix} X \\ I_k \end{pmatrix}$

Related: Burer and Monteiro [2003]

Recap

- **Quadratic matrix programs**

Recap

- **Quadratic matrix programs**
  - Easy if (**Lagr.**) dual has optimal solution γ^* with $A(\gamma^*) \succ 0$

$$X^* = \arg\min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma :\, A(\gamma) \succeq 0} q(\gamma, X)$$

Recap

- **Quadratic matrix programs**
  - Easy if (**Lagr.**) dual has optimal solution γ^* with $A(\gamma^*) \succ 0$

$$X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

- **Semidefinite programs**

Recap

- **Quadratic matrix programs**
  - Easy if (**Lagr.**) dual has optimal solution γ^* with $A(\gamma^*) \succ 0$

$$X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

- **Semidefinite programs**
  - k-exact if can be rewritten as easy QMP

Recap

- **Quadratic matrix programs**
  - Easy if (**Lagr.**) dual has optimal solution γ^* with $A(\gamma^*) \succ 0$

$$X^* = \arg\min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X)$$

- **Semidefinite programs**
  - k-exact if can be rewritten as easy QMP
  - Easy QMP reformulation $\approx$ Burer–Monteiro with symmetry removed

Recap

- **Quadratic matrix programs**
  - Easy if (**Lagr.**) dual has optimal solution γ^* with $A(\gamma^*) \succ 0$

$$X^* = \arg\min_{X \in \mathbb{R}^{n \times k}} \sup_{\gamma :\, A(\gamma) \succeq 0} q(\gamma, X)$$

- **Semidefinite programs**
  - k-exact if can be rewritten as easy QMP
  - Easy QMP reformulation $\approx$ Burer–Monteiro with symmetry removed
- **Next**: A new FOM for easy QMPs

Deriving a strongly convex minimax problem

$$A(\gamma^*) \succ 0 \qquad \text{and} \qquad X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

Deriving a strongly convex minimax problem

$$A(\gamma^*) \succ 0 \qquad \text{and} \qquad X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

- **Thought experiment**: By strict compl., X^* is opt. of strongly conv. function

$$X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} q(\gamma^*, X)$$

Deriving a strongly convex minimax problem

$$A(\gamma^*) \succ 0 \qquad \text{and} \qquad X^* = \operatorname*{arg\,min}_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

- **Thought experiment**: By strict compl., X^* is opt. of strongly conv. function

$$X^* = \operatorname*{arg\,min}_{X \in \mathbb{R}^{n\times k}} q(\gamma^*, X)$$

Theorem (Certificate of strict compl. gives strongly conv. reform.)

Suppose $\gamma^* \in \mathcal{C} \subseteq \mathbb{R}^m$ and $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$, then

$$X^* = \operatorname*{arg\,min}_{X \in \mathbb{R}^{n\times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

Deriving a strongly convex minimax problem

$$A(\gamma^*) \succ 0 \qquad \text{and} \qquad X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \sup_{\gamma:\, A(\gamma)\succeq 0} q(\gamma, X)$$

- **Thought experiment**: By strict compl., X^* is opt. of strongly conv. function

$$X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} q(\gamma^*, X)$$

Theorem (Certificate of strict compl. gives strongly conv. reform.)

Suppose $\gamma^* \in \mathcal{C} \subseteq \mathbb{R}^m$ and $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$, then

$$X^* = \arg\min_{X \in \mathbb{R}^{n\times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

- Strongly convex in X, no hidden PSD constraint

Deriving a strongly convex minimax problem

$$A(\gamma^*) \succ 0 \qquad \text{and} \qquad X^* = \underset{X \in \mathbb{R}^{n \times k}}{\arg\min} \sup_{\gamma:\, A(\gamma) \succeq 0} q(\gamma, X)$$

- **Thought experiment**: By strict compl., X^* is opt. of strongly conv. function

$$X^* = \underset{X \in \mathbb{R}^{n \times k}}{\arg\min}\, q(\gamma^*, X)$$

Theorem (Certificate of strict compl. gives strongly conv. reform.)

Suppose $\gamma^* \in \mathcal{C} \subseteq \mathbb{R}^m$ and $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$, then

$$X^* = \underset{X \in \mathbb{R}^{n \times k}}{\arg\min} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

- Strongly convex in X, no hidden PSD constraint
- **Alg. plan**: Construct $\mathcal{C}$, solve strongly convex reformulation

CautiousAGD

- How to solve strongly convex quadratic matrix minimax program?

$$(\textbf{QMMP}) = \min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

Related: Devolder et al. [2013, 2014], Nesterov [2005, 2018]

CautiousAGD

- How to solve strongly convex quadratic matrix minimax program?

$$(\textbf{QMMP}) = \min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

  - Accelerated gradient descent (AGD) method for minimax functions

---

Related: Devolder et al. [2013, 2014], Nesterov [2005, 2018]

- How to solve strongly convex quadratic matrix minimax program?

$$(\textbf{QMMP}) = \min_{X \in \mathbb{R}^{n\times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

  - Accelerated gradient descent (AGD) method for minimax functions
  - **Issue**: Requires solving the following prox-map in each iteration

$$X_{t+1} = \arg\min_{X \in \mathbb{R}^{n\times k}} \max_{\gamma \in \mathcal{C}} \left(\frac{L}{2} \left\| X - \hat{X}_t \right\|_F^2 + \left\langle \nabla q(\gamma, \hat{X}_t), X - \hat{X}_t \right\rangle + q(\gamma, \hat{X}_t) \right)$$

Related: Devolder et al. [2013, 2014], Nesterov [2005, 2018]

CautiousAGD

- How to solve strongly convex quadratic matrix minimax program?

$$(\textbf{QMMP}) = \min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

  - Accelerated gradient descent (AGD) method for minimax functions
  - **Issue**: Requires solving the following prox-map in each iteration

$$X_{t+1} = \arg\min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} \left(\frac{L}{2} \left\| X - \hat{X}_t \right\|_F^2 + \left\langle \nabla q(\gamma, \hat{X}_t), X - \hat{X}_t \right\rangle + q(\gamma, \hat{X}_t) \right)$$

    - Solve prox-map approximately and bound error in AGD

Related: Devolder et al. [2013, 2014], Nesterov [2005, 2018]

CautiousAGD

- How to solve strongly convex quadratic matrix minimax program?

$$(\textbf{QMMP}) = \min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} q(\gamma, X)$$

  - Accelerated gradient descent (AGD) method for minimax functions
  - **Issue**: Requires solving the following prox-map in each iteration

$$X_{t+1} = \arg\min_{X \in \mathbb{R}^{n \times k}} \max_{\gamma \in \mathcal{C}} \left(\frac{L}{2} \left\| X - \hat{X}_t \right\|_F^2 + \left\langle \nabla q(\gamma, \hat{X}_t), X - \hat{X}_t \right\rangle + q(\gamma, \hat{X}_t) \right)$$

    - Solve prox-map approximately and bound error in AGD
  - $\longrightarrow$ CautiousAGD

Related: Devolder et al. [2013, 2014], Nesterov [2005, 2018]

CautiousAGD guarantees

Theorem (CautiousAGD)

CautiousAGD produces iterates X_t such that

$$\max_{\gamma\in\mathcal{C}} q(\gamma, X_t) \leq \min_{X} \max_{\gamma\in\mathcal{C}} q(\gamma, X) + \epsilon$$

after $O\left(\log(\epsilon^{-1})\right)$ iterations, $O(m\epsilon^{-1/2})$ matrix-vector products per iteration

CertSDP

- How to construct $\mathcal{C}$?

CertSDP

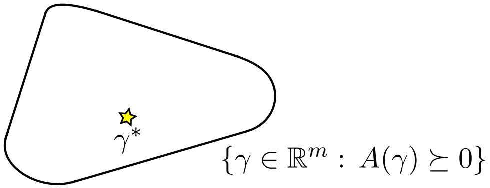
- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$

CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$

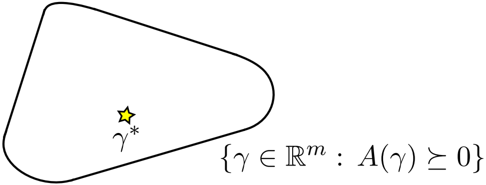
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$



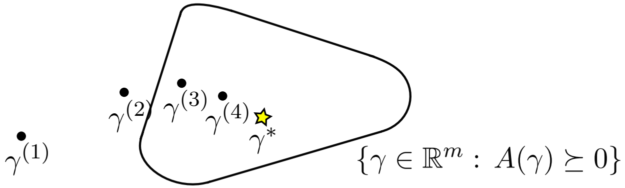
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$



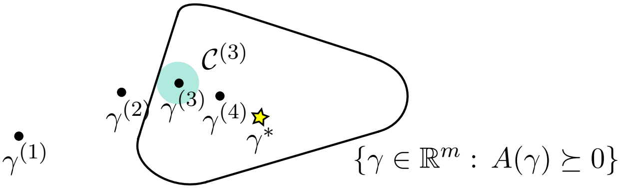
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$

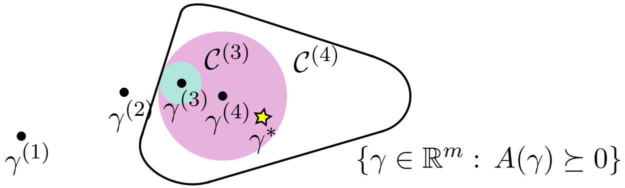


CertSDP

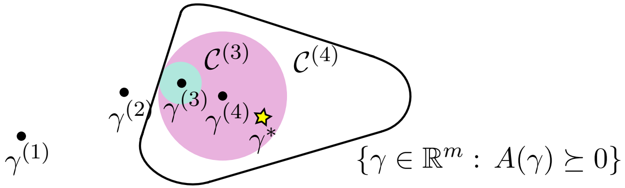
- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$



- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$

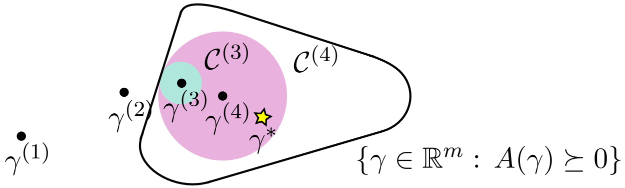


- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$
  - **Issue**: How do we know close enough?



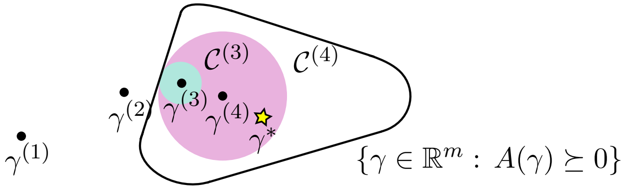
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$
  - **Issue**: How do we know close enough?
    Use CautiousAGD!



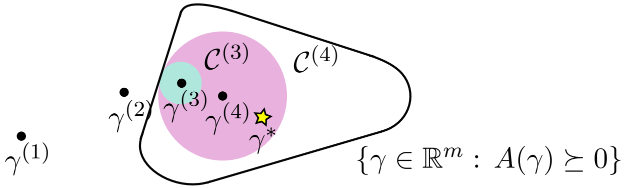
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$
  - **Issue**: How do we know close enough?
    Use CautiousAGD!
    - If $\gamma^* \in \mathcal{C}^{(i)}$ then CautiousAGD converges to X^* rapidly



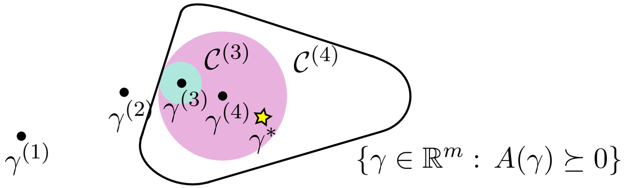
CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$
  - **Issue**: How do we know close enough?
    Use CautiousAGD!
    - If $\gamma^* \in \mathcal{C}^{(i)}$ then CautiousAGD converges to X^* rapidly
    - Monitor convergence!

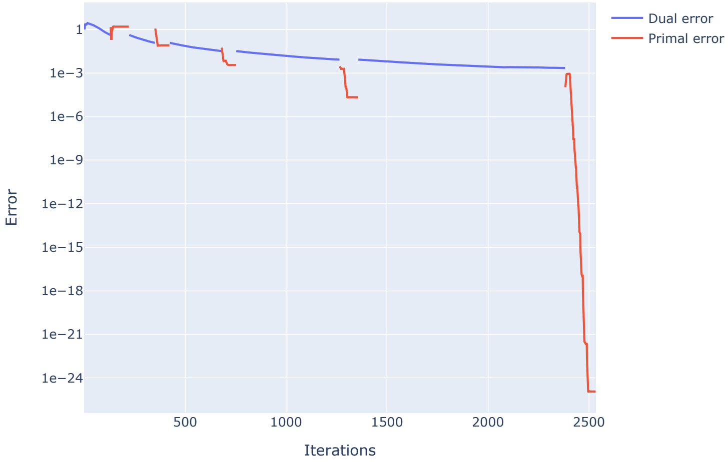


CertSDP

- How to construct $\mathcal{C}$? $\gamma^* \in \mathcal{C}$, $A(\gamma) \succ 0$ for all $\gamma \in \mathcal{C}$
  - γ^* is optimizer of dual problem, there exist (slow) algorithms $\gamma^{(i)} \to \gamma^*$
  - **Issue**: How do we know close enough?
    Use CautiousAGD!
    - If $\gamma^* \in \mathcal{C}^{(i)}$ then CautiousAGD converges to X^* rapidly
    - Monitor convergence!
  - $\longrightarrow$ CertSDP



CertSDP convergence behavior



CertSDP guarantees

Theorem (CertSDP)

CertSDP produces iterates X_t such that

$$\left\langle M_0, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle \leq \mathrm{Opt}_{\mathsf{SDP}} + \epsilon \qquad \left\| \left(\left\langle M_i, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle + d_i \right)_i \right\|_2 \leq \epsilon$$

Related: Ding et al. [2021], Friedlander and Macêdo [2016], Shinde et al. [2021], Yurtsever et al. [2021]

CertSDP guarantees

Theorem (CertSDP)

CertSDP produces iterates X_t such that

$$\left\langle M_0, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle \leq \mathrm{Opt}_{\mathsf{SDP}} + \epsilon \qquad \left\| \left(\left\langle M_i, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle + d_i \right)_i \right\|_2 \leq \epsilon$$

- **Iteration count**: $O(1) + O\left(\log(\epsilon^{-1})\right)$

Related: Ding et al. [2021], Friedlander and Macêdo [2016], Shinde et al. [2021], Yurtsever et al. [2021]

CertSDP guarantees

Theorem (CertSDP)

CertSDP produces iterates X_t such that

$$\left\langle M_0, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle \leq \text{Opt}_{\mathsf{SDP}} + \epsilon \qquad \left\| \left(\left\langle M_i, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle + d_i \right)_i \right\|_2 \leq \epsilon$$

- **Iteration count**: $O(1) + O\left(\log(\epsilon^{-1})\right)$
- **Iteration complexity**: $O(m\epsilon^{-1})$ matrix-vector products per iteration

---

Related: Ding et al. [2021], Friedlander and Macêdo [2016], Shinde et al. [2021], Yurtsever et al. [2021]

Theorem (CertSDP)

CertSDP produces iterates X_t such that

$$\left\langle M_0, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle \leq \mathrm{Opt}_{\mathsf{SDP}} + \epsilon \qquad \left\| \left(\left\langle M_i, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle + d_i \right)_i \right\|_2 \leq \epsilon$$

- **Iteration count**: $O(1) + O\left(\log(\epsilon^{-1})\right)$
- **Iteration complexity**: $O(m\epsilon^{-1})$ matrix-vector products per iteration
- **Storage**: $O(nk + m)$ storage

Related: Ding et al. [2021], Friedlander and Macêdo [2016], Shinde et al. [2021], Yurtsever et al. [2021]

CertSDP guarantees

Theorem (CertSDP)

CertSDP produces iterates X_t such that

$$\left\langle M_0, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle \leq \mathrm{Opt}_{\mathsf{SDP}} + \epsilon \qquad \left\| \left(\left\langle M_i, \begin{pmatrix} X_t X_t^\intercal & X_t \\ X_t^\intercal & I_k \end{pmatrix} \right\rangle + d_i \right)_i \right\|_2 \leq \epsilon$$

- **Iteration count**: $O(1) + O\left(\log(\epsilon^{-1})\right)$
- **Iteration complexity**: $O(m\epsilon^{-1})$ matrix-vector products per iteration
- **Storage**: $O(nk + m)$ storage

- Compare: $O(nk + m)$ storage, $O(\epsilon^{-2})$ iterations, $O(m\epsilon^{-1/2})$ per iteration

Related: Ding et al. [2021], Friedlander and Macêdo [2016], Shinde et al. [2021], Yurtsever et al. [2021]

Numerical results: experimental setup

- Random instances of **distance-minimization QMP**

$$\inf_{X\in\mathbb{R}^{n\times k}} \left\{ \|X\|_F^2 : q_i(X) = 0,\ \forall i \in [m] \right\}$$

with strict complementarity

Related: Ding et al. [2021], O'Donoghue et al. [2016], Souto et al. [2020], Yurtsever et al. [2021]

Numerical results: experimental setup

- Random instances of **distance-minimization QMP**

$$\inf_{X\in\mathbb{R}^{n\times k}} \left\{ \|X\|_F^2 : q_i(X) = 0, \forall i \in [m] \right\}$$

  with strict complementarity
- **Algorithms**: CertSDP, CSSDP, SketchyCGAL*, ProxSDP, SCS

Related: Ding et al. [2021], O'Donoghue et al. [2016], Souto et al. [2020], Yurtsever et al. [2021]

- Random instances of **distance-minimization QMP**

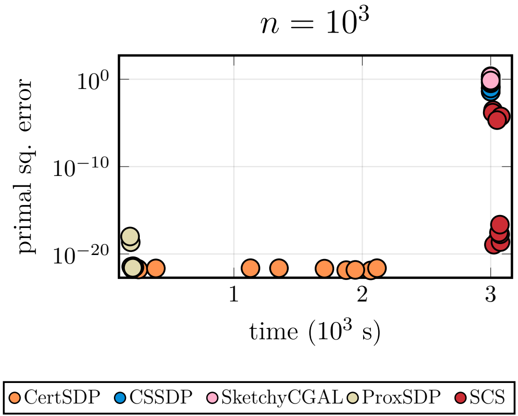
$$\inf_{X \in \mathbb{R}^{n \times k}} \left\{ \|X\|_F^2 : q_i(X) = 0, \forall i \in [m] \right\}$$

  with strict complementarity

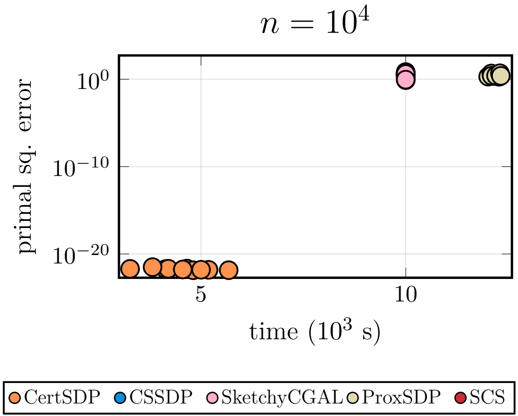
- **Algorithms**: CertSDP, CSSDP, SketchyCGAL*, ProxSDP, SCS
- $k = m = 10$, $n = 10^3, 10^4, 10^5$

Related: Ding et al. [2021], O'Donoghue et al. [2016], Souto et al. [2020], Yurtsever et al. [2021]

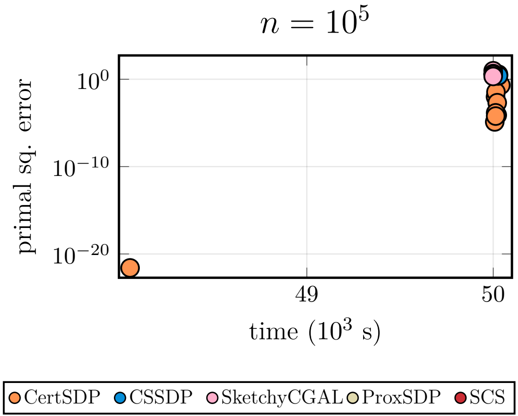
Numerical results: convergence comparisons



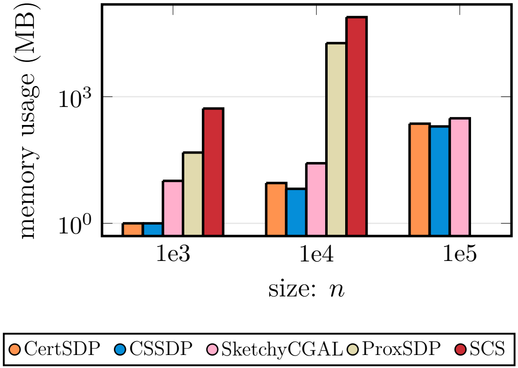
Numerical results: convergence comparisons



Numerical results: convergence comparisons



Numerical results: memory usage



Conclusion

- **Summary**: New FOM for k-exact SDPs

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement
  - Promising numerical performance as well!

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement
  - Promising numerical performance as well!
- **Questions**:

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement
  - Promising numerical performance as well!
- **Questions**:
  - Rates depend polynomially on $\lambda_{\max}(A(\gamma^*))$, $\lambda_{\min}(A(\gamma^*))$.
    What are the optimal rates? Lower bounds?

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement
  - Promising numerical performance as well!
- **Questions**:
  - Rates depend polynomially on $\lambda_{\max}(A(\gamma^*))$, $\lambda_{\min}(A(\gamma^*))$.
    What are the optimal rates? Lower bounds?
  - Understanding how error propagates if Y_W^* only known up to some error?

Conclusion

- **Summary**: New FOM for k-exact SDPs
  - Low iteration complexity, cheap iterations, and low storage requirement
  - Promising numerical performance as well!
- **Questions**:
  - Rates depend polynomially on $\lambda_{\max}(A(\gamma^*))$, $\lambda_{\min}(A(\gamma^*))$.
    What are the optimal rates? Lower bounds?
  - Understanding how error propagates if Y_W^* only known up to some error?

Thank you! Questions?

`https://arxiv.org/abs/2206.00224`

Alizadeh, F., Haeberly, J. A., and Overton, M. L. (1997). Complementarity and nondegeneracy in semidefinite programming. *Math. Program.*, 77:111–128.

Beck, A. (2007). Quadratic matrix programming. *SIAM J. Optim.*, 17(4):1224–1238.

Beck, A., Drori, Y., and Teboulle, M. (2012). A new semidefinite programming relaxation scheme for a class of quadratic matrix problems. *Oper. Res. Lett.*, 40(4):298–302.

Burer, S. and Monteiro, R. D. (2003). A nonlinear programming algorithm for solving semidefinite programs via low-rank factorization. *Math. Program.*, 95:329–357.

Devolder, O., Glineur, F., and Nesterov, Y. (2013). First-order methods with inexact oracle: the strongly convex case. Technical Report 2013016.

Devolder, O., Glineur, F., and Nesterov, Y. (2014). First-order methods of smooth convex optimization with inexact oracle. *Math. Program.*, 146(1):37–75.

Ding, L., Yurtsever, A., Cevher, V., Tropp, J. A., and Udell, M. (2021). An optimal-storage approach to semidefinite programming using approximate complementarity. *SIAM J. Optim.*, 31(4):2695–2725.

Friedlander, M. P. and Macêdo, I. (2016). Low-rank spectral optimization via gauge duality. *SIAM Journal on Scientific Computing*, 38(3):A1616–A1638.

References II

Nesterov, Y. (2005). Excessive gap technique in nonsmooth convex minimization. *SIAM J. Optim.*, 16(1):235–249.

Nesterov, Y. (2018). *Lectures on convex optimization*. Number 137 in Springer Optim. and its Appl. 2 edition.

O'Donoghue, B., Chu, E., Parikh, N., and Boyd, S. (2016). Conic optimization via operator splitting and homogeneous self-dual embedding. *Journal of Optimization Theory and Applications*, 169(3):1042–1068.

Shinde, N., Narayanan, V., and Saunderson, J. (2021). Memory-efficient structured convex optimization via extreme point sampling. *SIAM Journal on Mathematics of Data Science*, 3(3):787–814.

Souto, M., Garcia, J. D., and Veiga, A. (2020). Exploiting low-rank structure in semidefinite programming by approximate operator splitting. *Optimization*, pages 1–28.

Yurtsever, A., Tropp, J. A., Fercoq, O., Udell, M., and Cevher, V. (2021). Scalable semidefinite programming. *SIAM Journal on Mathematics of Data Science*, 3(1):171–200.