

Dimensionality reduction of SDPs through sketching

Daniel Stilck França

Technische Universität München



Workshop on "Probabilistic techniques and Quantum Information Theory", Institut Henri Poincaré

Joint work with Andreas Bluhm

arXiv:1707.09863

Semidefinite Programs (SDPs)

Semidefinite programs are constrained optimization problems of the form:

$$\begin{aligned} & \text{maximize} && \text{tr}(AX) \\ & \text{subject to} && \text{tr}(B_i X) \leq \gamma_i, \quad i \in [m] \\ & && X \geq 0, \end{aligned}$$

where $A, B_1, \dots, B_m \in \mathcal{M}_D^{\text{sym}}$ are symmetric matrices and $\gamma_1, \dots, \gamma_m \in \mathbb{R}$.

Semidefinite Programs (SDPs)

Semidefinite programs are constrained optimization problems of the form:

$$\begin{aligned} & \text{maximize} && \text{tr}(AX) \\ & \text{subject to} && \text{tr}(B_i X) \leq \gamma_i, \quad i \in [m] \\ & && X \geq 0, \end{aligned}$$

where $A, B_1, \dots, B_m \in \mathcal{M}_D^{\text{sym}}$ are symmetric matrices and $\gamma_1, \dots, \gamma_m \in \mathbb{R}$. They can be seen as a generalization of linear programs and have many applications throughout QIT.

Semidefinite Programs (SDPs)

Semidefinite programs are constrained optimization problems of the form:

$$\begin{aligned} & \text{maximize} && \text{tr}(AX) \\ & \text{subject to} && \text{tr}(B_i X) \leq \gamma_i, \quad i \in [m] \\ & && X \geq 0, \end{aligned}$$

where $A, B_1, \dots, B_m \in \mathcal{M}_D^{\text{sym}}$ are symmetric matrices and $\gamma_1, \dots, \gamma_m \in \mathbb{R}$. They can be seen as a generalization of linear programs and have many applications throughout QIT. They can be written in many equivalent forms. Will call this one *sketchable* SDP. Any SDP can be formulated in this form.

Semidefinite Programs (SDPs)

- **Good news:** in most cases they can be solved in polynomial time!

Semidefinite Programs (SDPs)

- **Good news:** in most cases they can be solved in polynomial time!
- Using the ellipsoid method we can solve them in $\mathcal{O}(\max\{m, D^2\} D^6 \log(1/\zeta))$ time, where ζ is the error tolerance.

Semidefinite Programs (SDPs)

- **Good news:** in most cases they can be solved in polynomial time!
- Using the ellipsoid method we can solve them in $\mathcal{O}(\max\{m, D^2\} D^6 \log(1/\zeta))$ time, where ζ is the error tolerance.
- **Bad news:** scaling still prohibitive for high dimensional problems! Especially when it comes to memory. Try running an SDP with $D \sim 10^3$ on your laptop and you will already run out of memory.

Semidefinite Programs (SDPs)

- **Good news:** in most cases they can be solved in polynomial time!
- Using the ellipsoid method we can solve them in $\mathcal{O}(\max\{m, D^2\} D^6 \log(1/\zeta))$ time, where ζ is the error tolerance.
- **Bad news:** scaling still prohibitive for high dimensional problems! Especially when it comes to memory. Try running an SDP with $D \sim 10^3$ on your laptop and you will already run out of memory.
- Need techniques to solve larger problems. Ideally using available solvers.

- Apply a positive linear map $\Phi : \mathcal{M}_D \rightarrow \mathcal{M}_d$ to constraints s.t. $\text{tr}(\Phi(B_i)\Phi(X^*)) \approx \text{tr}(B_i X^*)$ holds with high probability. Here X^* is a solution of the SDP.

- Apply a positive linear map $\Phi : \mathcal{M}_D \rightarrow \mathcal{M}_d$ to constraints s.t. $\text{tr}(\Phi(B_i)\Phi(X^*)) \approx \text{tr}(B_i X^*)$ holds with high probability. Here X^* is a solution of the SDP.
- Solve the SDP defined by $\Phi(B_i)$ and show that its value is not far from the value of the original problem.

- Apply a positive linear map $\Phi : \mathcal{M}_D \rightarrow \mathcal{M}_d$ to constraints s.t. $\text{tr}(\Phi(B_i)\Phi(X^*)) \approx \text{tr}(B_i X^*)$ holds with high probability. Here X^* is a solution of the SDP.
- Solve the SDP defined by $\Phi(B_i)$ and show that its value is not far from the value of the original problem.
- If $d \ll D$ and computing $\Phi(B_i)$ is cheap, then this gives a computational advantage.

Theorem (SDPs cannot be sketched)

Let $\Phi : \mathcal{M}_{2D} \rightarrow \mathbb{R}^d$ be a random linear map such that for all sketchable SDPs there exists an algorithm which allows us to estimate the value of an SDP up to a constant factor $1 \leq \tau < \frac{2}{\sqrt{3}}$ given the sketch $\{\Phi(A), \Phi(B_1), \dots, \Phi(B_m)\}$ with probability at least $9/10$. Then $d = \Omega(D^2)$.

Theorem (Not all SDPs can be sketched)

Let $\Phi : \mathcal{M}_{2D} \rightarrow \mathbb{R}^d$ be a random linear map such that for all sketchable SDPs there exists an algorithm which allows us to estimate the value of an SDP up to a constant factor $1 \leq \tau < \frac{2}{\sqrt{3}}$ given the sketch $\{\Phi(A), \Phi(B_1), \dots, \Phi(B_m)\}$ with probability at least $9/10$. Then $d = \Omega(D^2)$.

Definition (Johnson-Lindenstrauss transform)

A random matrix $S \in \mathcal{M}_{d,D}$ is a Johnson-Lindenstrauss transform (JLT) with parameters (ϵ, δ, k) if with probability at least $1 - \delta$, for any k -element subset $V \subseteq \mathbb{K}^D$, for all $v, w \in V$ it holds that

$$|\langle Sv, Sw \rangle - \langle v, w \rangle| \leq \epsilon \|v\|_2 \|w\|_2.$$

Definition (Johnson-Lindenstrauss transform)

A random matrix $S \in \mathcal{M}_{d,D}$ is a Johnson-Lindenstrauss transform (JLT) with parameters (ϵ, δ, k) if with probability at least $1 - \delta$, for any k -element subset $V \subseteq \mathbb{K}^D$, for all $v, w \in V$ it holds that

$$|\langle Sv, Sw \rangle - \langle v, w \rangle| \leq \epsilon \|v\|_2 \|w\|_2.$$

- Example: $S = \frac{1}{\sqrt{d}}R \in \mathcal{M}_{d,D}$, where the entries of R are i.i.d. standard Gaussian random variables. If $d = \Omega(\epsilon^{-2} \log(k\delta^{-1}))$, then S is an (ϵ, δ, k) -JLT.

Lemma (Sketching Hilbert-Schmidt Scalar Product)

Let $B_1, \dots, B_m \in \mathcal{M}_D$ and $S \in \mathcal{M}_{d,D}$ be an (ϵ, δ, k) -JLT with $\epsilon \leq 1$ and k such that

$$k \geq \sum_{i=1}^m \text{rank}(B_i).$$

Then, with probability at least $1 - \delta$,

$$\forall i, j \in [m] : |\text{tr}(SB_i S^T SB_j S^T) - \text{tr}(B_i B_j)| \leq 3\epsilon \|B_i\|_1 \|B_j\|_1$$

$$|\operatorname{tr}(SB_iS^T SB_jS^T) - \operatorname{tr}(B_iB_j)| \leq 3\epsilon \|B_i\|_1 \|B_j\|_1$$

- Scaling with $\|\cdot\|_1$ is undesirable. Normal JLT gives scaling with $\|\cdot\|_2$.

$$|\operatorname{tr}(SB_iS^T SB_jS^T) - \operatorname{tr}(B_iB_j)| \leq 3\epsilon \|B_i\|_1 \|B_j\|_1$$

- Scaling with $\|\cdot\|_1$ is undesirable. Normal JLT gives scaling with $\|\cdot\|_2$.
- Proof of the inequality admittedly crude. Can we improve the inequality?

Theorem (No Johnson-Lindenstrauss with positive maps)

Let $\Phi : \mathcal{M}_D \rightarrow \mathcal{M}_d$ be a random positive map such that with strictly positive probability for any $Y_1, \dots, Y_{D+1} \in \mathcal{M}_D$ and $0 < \epsilon < \frac{1}{4}$ we have

$$|\operatorname{tr}(\Phi(Y_i)^T \Phi(Y_j)) - \operatorname{tr}(Y_i^T Y_j)| \leq \epsilon \|Y_i\|_2 \|Y_j\|_2.$$

Then $d = \Omega(D)$.

- **Assumptions:** uniform bounds on $\|A\|_1, \|B_1\|_1, \dots, \|B_m\|_1$ and $\|X^*\|_1$, where X^* is an optimal point of the SDP. Standard regularity assumptions on SDP.

- **Assumptions:** uniform bounds on $\|A\|_1, \|B_1\|_1, \dots, \|B_m\|_1$ and $\|X^*\|_1$, where X^* is an optimal point of the SDP. Standard regularity assumptions on SDP.

Consider the sketchable SDP of dimension D :

$$\begin{array}{ll} \text{maximize} & \text{tr}(AX) \\ \text{subject to} & \text{tr}(B_i X) \leq \gamma_i, \quad i \in [m] \\ & X \geq 0. \end{array}$$

The Algorithm

Now pick a (ϵ, δ, k) JL-transform $S \in \mathcal{M}_{d,D}$, where

$$k \geq \text{rank}(X^*) + \text{rank}(A) + \sum_{i=1}^m \text{rank}(B_i),$$

and consider the SDP of dimension d :

$$\begin{aligned} & \text{maximize} && \text{tr}(SAS^T Y) \\ & \text{subject to} && \text{tr}(SB_i S^T Y) \leq \gamma_i, \quad i \in [m] \\ & && Y \geq 0. \end{aligned}$$

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0 \end{aligned}$$

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0 \end{aligned}$$

Call this SDP the sketched SDP!

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0 \end{aligned}$$

Call this SDP the sketched SDP! Solve this SDP!

The Algorithm

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0. \end{aligned}$$

- Bound on HS scalar product gives that SX^*S^T is a feasible point of the relaxed problem with probability $\geq 1 - \delta$.

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0. \end{aligned}$$

- Bound on HS scalar product gives that SX^*S^T is a feasible point of the relaxed problem with probability $\geq 1 - \delta$.
- We have $\text{tr} \left(SAS^T SX^*S^T \right) \geq \text{tr} \left(AX^* \right) - 3\epsilon \|X^*\|_1 \|A\|_1$.

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0. \end{aligned}$$

- Bound on HS scalar product gives that SX^*S^T is a feasible point of the relaxed problem with probability $\geq 1 - \delta$.
- We have $\text{tr} \left(SAS^T SX^*S^T \right) \geq \text{tr} \left(AX^* \right) - 3\epsilon \|X^*\|_1 \|A\|_1$.
- But $\text{tr} \left(AX^* \right) = \alpha$ is the value of the sketchable SDP! We therefore obtain $\alpha_S + 3\epsilon \|X^*\|_1 \|A\|_1 \geq \alpha$, where α_S is the value of the sketched SDP.

Relax the constraints:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0. \end{aligned}$$

- Bound on HS scalar product gives that SX^*S^T is a feasible point of the relaxed problem with probability $\geq 1 - \delta$.
- We have $\text{tr} \left(SAS^T SX^*S^T \right) \geq \text{tr} \left(AX^* \right) - 3\epsilon \|X^*\|_1 \|A\|_1$.
- But $\text{tr} \left(AX^* \right) = \alpha$ is the value of the sketchable SDP! We therefore obtain $\alpha_S + 3\epsilon \|X^*\|_1 \|A\|_1 \geq \alpha$, where α_S is the value of the sketched SDP.

Theorem

Let $A, B_1, \dots, B_m \in \mathcal{M}_D^{\text{sym}}$, $\eta, \gamma_1, \dots, \gamma_m \in \mathbb{R}$ and $\epsilon > 0$. Denote by α the value of the sketchable SDP and assume it is attained at an optimal point X^* which satisfies $\text{tr}(X^*) \leq \eta$. Moreover, let $S \in \mathcal{M}_{d,D}$ be an (ϵ, δ, k) -JLT, with

$$k \geq \text{rank}(X^*) + \text{rank}(A) + \sum_{i=1}^m \text{rank}(B_i).$$

Let α_S be the value of the sketched SDP defined by A , B_i and S . Then

$$\alpha_S + 3\epsilon\eta\|A\|_1 \geq \alpha$$

with probability at least $1 - \delta$.

- Can it be the case that $\alpha_S \gg \alpha$?

- Can it be the case that $\alpha_S \gg \alpha$?
- Depends on how stable your SDP is!

Let Y^* be an optimal point of your sketched SDP. That is, a solution of:

$$\begin{aligned} & \text{maximize} && \text{tr} \left(SAS^T Y \right) \\ & \text{subject to} && \text{tr} \left(SB_i S^T Y \right) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & && Y \geq 0 \end{aligned}$$

By the cyclicity of the trace, $S^T Y^* S$ is a feasible point of

$$\begin{array}{ll} \text{maximize} & \text{tr}(AX) \\ \text{subject to} & \text{tr}(B_i X) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & X \geq 0 \end{array}$$

with value α_S .

By the cyclicity of the trace, $S^T Y^* S$ is a feasible point of

$$\begin{array}{ll} \text{maximize} & \text{tr}(AX) \\ \text{subject to} & \text{tr}(B_i X) \leq \gamma_i + 3\epsilon \|B_i\|_1 \|X^*\|_1, \quad i \in [m] \\ & X \geq 0 \end{array}$$

with value α_S . This is just a perturbed version of the original SDP!

Theorem (Lower Bound in terms of α_S)

For a sketchable SDP with $\gamma_i = 1$ and $\kappa = \max_{i \in [m]} \|B_i\|_1$, we have that

$$\frac{\alpha_S}{1 + \nu} \leq \alpha,$$

where $\nu = 3\epsilon\eta\kappa$. Moreover, denoting by X_S^* an optimal point of the sketched SDP, we have that $\frac{1}{1+\nu} S^T X_S^* S$ is a feasible point of the sketchable SDP that attains this lower bound.

Theorem

For a sketchable SDP with $\gamma_i = 1$ and $\kappa = \max_{i \in [m]} \|B_i\|_1$, we have that

$$\frac{\alpha_S}{1 + \nu} \leq \alpha \leq \alpha_S + 3\epsilon\eta\|A\|_1,$$

where $\nu = 3\epsilon\eta\kappa$.

- Assuming $\|A\|_1, \|B_1\|, \dots, \|B_m\|_1, \|X^*\|_1 = \mathcal{O}(1)$ and ϵ, δ, ζ fixed we obtain:

Theorem

Let $A, B_1, \dots, B_m \in \mathcal{M}_D^{\text{sym}}$, of a sketchable SDP be given. Furthermore, let $\text{SDP}(m, d)$ be the complexity of solving a sketchable SDP of dimension d and m constraints up to some given precision. Then a number of

$$\mathcal{O}(D^2 m \log(k)) + \text{SDP}(m, \log(k))$$

operations is needed to generate and solve the sketched SDP, where $k \leq (m + 2)D^2$ is defined as before.

Complexity and Memory Considerations

- Assuming ϵ, δ fixed, sketching gives a speedup as long as the complexity of solving the SDP directly is $\Omega(mD^{2+\mu})$, where $\mu > 0$.

Complexity and Memory Considerations

- Assuming ϵ, δ fixed, sketching gives a speedup as long as the complexity of solving the SDP directly is $\Omega(mD^{2+\mu})$, where $\mu > 0$.
- Need only to store $\mathcal{O}(m\epsilon^{-4} \log(mk/\delta)^2)$ entries to solve the sketched problem.

Given observables $A, B \in \mathcal{M}_D^{\text{sym}}$, consider uncertainty relations of the form

$$\text{tr}(A^2\rho) + \text{tr}(B^2\rho) \geq c$$

for all states ρ s.t. $\text{tr}(A\rho) \in (a - \epsilon, a + \epsilon)$ and $\text{tr}(B\rho) \in (b - \epsilon, b + \epsilon)$.

Given observables $A, B \in \mathcal{M}_D^{\text{sym}}$, consider uncertainty relations of the form

$$\text{tr}(A^2\rho) + \text{tr}(B^2\rho) \geq c$$

for all states ρ s.t. $\text{tr}(A\rho) \in (a - \epsilon, a + \epsilon)$ and $\text{tr}(B\rho) \in (b - \epsilon, b + \epsilon)$. Finding the optimal c can easily be cast as an SDP.

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & \text{tr}(X) = 1, \\ & X \geq 0. \end{array}$$

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & \text{tr}(X) = 1, \\ & X \geq 0. \end{array}$$

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & \text{tr}(X) = 1, \\ & X \geq 0. \end{array}$$

- Can't handle $\text{tr}(X) = 1$ as a constraint. Relax problem and drop it.

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & X \geq 0. \end{array}$$

- Can't handle $\text{tr}(X) = 1$ as a constraint. Relax problem and drop it.

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & X \geq 0. \end{array}$$

- Can't handle $\text{tr}(X) = 1$ as a constraint. Relax problem and drop it.
- If $\|B\|_1, \|A\|_1 = \mathcal{O}(1)$ and their nonzero spectrum "flat", we can show $\|X^*\|_1 = \mathcal{O}(1)$.

$$\begin{array}{ll} \text{minimize} & \text{tr}((A^2 + B^2)X) \\ \text{subject to} & \text{tr}(AX) \leq a \pm \epsilon, \\ & \text{tr}(BX) \leq b \pm \epsilon, \\ & X \geq 0. \end{array}$$

- Can't handle $\text{tr}(X) = 1$ as a constraint. Relax problem and drop it.
- If $\|B\|_1, \|A\|_1 = \mathcal{O}(1)$ and their nonzero spectrum "flat", we can show $\|X^*\|_1 = \mathcal{O}(1)$.
- Example: A, B of fixed rank and nonzero spectrum contained in some compact interval.

D	d	Value	Error L.B.	M.R.T. Sketchable [s]	M.R.T Sketch [s]
200	50	0.0928	0.0429	6.73	0.663
200	100	0.0897	0.0401	6.51	1.336
500	100	0.0353	0.0181	96.5	1.35
500	200	0.0364	0.0152	96.4	6.81

Table: For each combination of the sketchable dimension (D) and dimension of the sketch (d) we have generated 40 instances of the uncertainty relation SDP. Here “M.R.T.” stands for mean running time, “L.B.” stands for lower bound we obtained from the sketch. The column “Value” stands for the optimal value of the sketchable SDP.

Thanks!