On the central path of semidefinite optimization
Degree and worst-case convergence rate

Ali Mohammad-Nezhad
Joint work with Saugata Basu

Workshop on Real Algebraic Geometry and Algorithms for Geometric Constraint Systems

June 18, 2021

PURDUE UNIVERSITY™

Overview

- Part 1: *Semidefinite optimization*
- Part 2: *Central path*
- Part 3: *Convergence rate and degree*

Part 1: Semidefinite optimization

Let $A^1, \ldots, A^m \in \mathbb{S}^n$, $C \in \mathbb{S}^n$, and $b := (b_1, \ldots, b_m)^T \in \mathbb{R}^m$.
$\mathbb{S}^n$: the vector space of symmetric $n \times n$ real matrices.

A semidefinite optimization (SDO) problem in primal-dual form:

$$(\mathrm{P}) \inf_X \left\{ \langle C, X \rangle \mid \langle A^i, X \rangle = b_i, \quad i = 1, \ldots, m, \ X \succeq 0 \right\},$$

$$(\mathrm{D}) \sup_{(y,S)} \left\{ b^T y \mid \sum_{i=1}^m y_i A^i + S = C, \ S \succeq 0 \right\}.$$

$\langle X, S \rangle = \mathrm{tr}(XS), \qquad \forall X, S \in \mathbb{S}^n.$

- We assume that all the coefficients $(\mathbf{A}, b, C)$ belong to $\mathbb{Z}$.

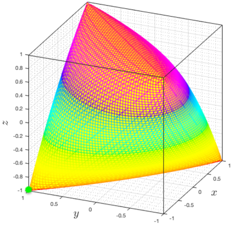
A SDO example

Minimization of $4x - 4y - 2z$ over the 3-elliptope:

$$\min\left\{4x - 4y - 2z \mid \begin{pmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{pmatrix} \succeq 0\right\}.$$

A unique optimal solution exists ($X^* + S^* \not\succ 0$):

$$X^* = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}, \quad y^* = (-4,\ -1,\ -1)^T, \quad S^* = \begin{pmatrix} 4 & 2 & -2 \\ 2 & 1 & -1 \\ -2 & -1 & 1 \end{pmatrix}.$$



PURDUE
UNIVERSITY

We assume that

- A^i for $i = 1, \ldots, m$ are linearly independent.
- $\exists$ a feasible $(X^\circ, y^\circ, S^\circ)$ s.t. $X^\circ, S^\circ \succ 0$.

Let $\mathrm{Sol(P)}$ and $\mathrm{Sol(D)}$ denote the primal and dual optimal sets.

- $\mathrm{Sol(P)} \times \mathrm{Sol(D)}$ is equal to all (X, y, S) s.t.

$$\begin{aligned} \langle A^i, X\rangle = b_i, \quad i = 1, \ldots, m, & \qquad X \succeq 0, \\ \sum_{i=1}^m y_i A^i + S = C, & \qquad S \succeq 0, \\ XS = 0. & \end{aligned}$$

- $\mathrm{Sol(P)}$ and $\mathrm{Sol(D)}$ are both nonempty and compact.

How to solve $(\mathrm{P})-(\mathrm{D})$?

Algebraically (Basu, Pollack, and Roy (BPR), Algorithm 14.9)

- $\{\langle A^i, X\rangle = b_i,\ i=1,\dots,m,\ X \succeq 0\}$ is a basic semi-algebraic set.
- $2^n - 1 + m$ polynomials of degree n.

Numerically (Interior point methods (IPMs))

- Iteratively solves a linear approximation of the *central path*:

$$\begin{aligned} \langle A^i, X\rangle &= b_i, \quad i=1,\dots,m, \\ \sum_{i=1}^{m} y_i A^i + S &= C, \\ XS + SX &= 2\mu I, \end{aligned}$$

while taking $\mu \downarrow 0$.

- Iteration complexity $O(\sqrt{n}L_\epsilon)$ for an ϵ-approximate solution.

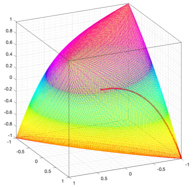
Trajectory of IPMs solutions

3-elliptope example:

$$X(\mu) = \begin{pmatrix} 1 & X_{12}(\mu) & -X_{12}(\mu) \\ X_{12}(\mu) & 1 & -2X_{12}^2(\mu) + \frac{\mu}{2}X_{12}(\mu) + 1 \\ -X_{12}(\mu) & -2X_{12}^2(\mu) + \frac{\mu}{2}X_{12}(\mu) + 1 & 1 \end{pmatrix}$$

- $X_{12}(\mu)$ is a real root of $2T^3 + (2 - \mu/2)T^2 - (\mu + 2)T - 2 = 0$.

Figure: IPMs follow the *central path* to solve $(\mathrm{P}) - (\mathrm{D})$.



Part 2: Central path

Central path(CP)

CP is a smooth semi-algebraic path $\phi(\mu):(0,\infty) \to \mathbb{S}^n \times \mathbb{R}^m \times \mathbb{S}^n$

$$\mathcal{C} := \Big\{(X,y,S,\mu) \mid \quad \langle A^i, X\rangle = b_i, \quad i = 1,\dots,m, \quad \sum_{i=1}^m y_i A^i + S = C, \quad XS + SX - 2\mu I_n = 0, \quad X, S \succ 0\Big\}.$$

- $\phi(\mu) = (X(\mu), y(\mu), S(\mu))$ is called a *central solution*.
- For every positive $\bar{\mu}$, $\phi\mid_{(0,\bar{\mu}]}$ is bounded.

Using real algebraic geometry:

Theorem (BPR, Prop. 3.18)

$\phi(\mu)$ can be continuously extended to $\mu = 0$, i.e., $\lim_{\mu\downarrow 0}\phi(\mu)$ exists.

Distance from the limit point (special case)

Let (X^{**}, y^{**}, S^{**}) be the limit point of CP.

If $X^{**} + S^{**} \succ 0$ (*strict complementarity*):

Theorem (Sturm et al., 2001)

(X^{**}, y^{**}, S^{**}) is the analytic center of the optimal set. Furthermore,

$$\|X(\mu) - X^{**}\| = O(\mu) \quad \textit{and} \quad \|S(\mu) - S^{**}\| = O(\mu).$$

- IPMs are superlinearly convergent under this condition.

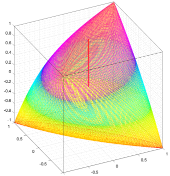


Figure: $\min\left\{ -2z \mid \begin{pmatrix} 1 & x & y \\ x & 1 & z \\ y & z & 1 \end{pmatrix} \succeq 0 \right\}.$

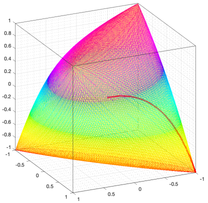
PURDUE UNIVERSITY

Distance from the limit point (general case)

How about when $X^{**} + S^{**} \not\succ 0$?

- IPMs convergence is no better than linear.
- Unknown worst-case convergence rate.
- The Lipschitzian bounds in previous case fails to exist.

The 3-elliptope example (using Newton-Puiseux theorem):



$\|X(\mu) - X^{**}\| = O(\sqrt{\mu})$ and $\|S(\mu) - S^{**}\| = O(\sqrt{\mu})$.

What happens if $X^{**} + S^{**} \not\succ 0$?

We already know that:

Theorem (M and Terlaky, 2019)

$$\operatorname{dist}\big(X(\mu), \operatorname{Sol(P)}\big) = O(\mu^{2^{1-n}}), \quad \operatorname{dist}\big((y(\mu), S(\mu)), \operatorname{Sol(D)}\big) = O(\mu^{2^{1-n}}).$$

A lower bound exists (De Klerk, 2001):

$$\max\left\{ -y_n \mid S = \begin{pmatrix} 1 & y_1 & y_2 & \dots & y_{n-1} \\ y_1 & y_2 & 0 & \dots & 0 \\ y_2 & 0 & y_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ y_{n-1} & 0 & \dots & 0 & y_n \end{pmatrix} \succeq 0 \right\}.$$

- $\|S(\mu) - S^{**}\| = \Omega(\mu^{2^{-n}})$.

Our goal here is to bound $\|X(\mu) - X^{**}\|$ and $\|S(\mu) - S^{**}\|$.

Part 3: Convergence rate and degree

Representation of CP (a semi-algebraic approach)

We identify symmetric matrices by vectors

$$X \mapsto \left(X_{11}, \sqrt{2}X_{12}, \ldots, \sqrt{2}X_{1n}, X_{22}, \sqrt{2}X_{23}, \ldots, \sqrt{2}X_{2n}, \ldots, X_{nn}\right)^T.$$

Then the CP system can be represented as

$$\mathcal{P} := \{\mathbf{A}x - b,\ \mathbf{A}^T y + s - c,\ (x \otimes_s I_n)s - 2\mu 1_{t(n)}\} \subset \mathbb{Z}[\mu, V].$$

$\mathbf{A} := \left(\text{svec}(A^1), \ldots, \text{svec}(A^m)\right)^T$;
$v := (x; y; s)$;
$\otimes_s$: the symmetric Kronecker product;
$t(n) := \binom{n}{2}$.

- A central solution is an isolated solution of $\text{Zer}(\mathcal{P}, \mathbb{R}^{m+2t(n)})$.

We compute real univariate representations describing $\mathrm{Zer}(\mathcal{P}, \mathbb{R}^{m+2t(n)})$.

Let $\mathcal{Q} \subset \mathbb{R}[X_1, \ldots, X_k]$ s.t. $\mathrm{Zer}(\mathcal{Q}, \mathbb{C}^k)$ is finite.

- A *real k-univariate representation* of $\hat{x} \in \mathrm{Zer}(\mathcal{Q}, \mathbb{R}^k)$ is (u, σ) with $u = \big(f, g_0, \ldots, g_k\big) \in \mathbb{R}[T]^{k+2}$ such that

$$\hat{x} = x_u(t_\sigma) := \left(\frac{g_1(t_\sigma)}{g_0(t_\sigma)}, \ldots, \frac{g_k(t_\sigma)}{g_0(t_\sigma)}\right) \in \mathbb{R}^k.$$

σ: the Thom encoding of a real root $t_\sigma \in \mathbb{R}$ of $f(T)$.
f and g_0 are coprime.

A bounded hypersurface

$\mathrm{Zer}(\mathcal{P}, \mathbb{R}^{m+2t(n)})$ may have an unbounded component. However,

$$\forall \bar{\mu} > 0 \;\; \exists \epsilon_{\bar{\mu}} \in \mathbb{Q}_+ \;\text{ s.t. }\; \phi(0, \bar{\mu}]) \subset \bigcup_{\mu} \mathrm{Zer}(\mathcal{P}, \mathbb{R}^{m+2t(n)}) \cap B(0, 1/\epsilon_{\bar{\mu}}).$$

Thus, we can define

$$Q = \|\mathbf{A}x - b\|^2 + \|\mathbf{A}^T y + s - c\|^2 + \|(x \otimes_s I_n)s - 2\mu 1_{t(n)}\|^2,$$
$$\tilde{Q} = Q^2 + (\epsilon_{\bar{\mu}}^2 (V_1^2 + \ldots + V_{m+2t(n)+1}^2) - 1)^2.$$

- $\mathrm{Zer}(\tilde{Q}, \mathbb{R}^{m+2t(n)+1})$ is the intersection of a ball and a cylinder based on $\mathrm{Zer}(\mathcal{P}, \mathbb{R}^{m+2t(n)})$.

Parameterized Bounded Algebraic Sampling (PBAS):

Theorem (Basu and M, 2021)

There exists $((f,g),\sigma)$ s.t. $f(\mu,t_\sigma)=0$, $g_0(\mu,t_\sigma)\neq 0$, *and* $\forall \mu \in (0,\bar{\mu}]$:

$$x_i(\mu) = \frac{g_i(\mu,t_\sigma)}{g_0(\mu,t_\sigma)} \in \mathbb{R}, \qquad i = 1\ldots,t(n),$$

$$y_i(\mu) = \frac{g_{t(n)+i}(\mu,t_\sigma)}{g_0(\mu,t_\sigma)} \in \mathbb{R}, \qquad i = 1\ldots,m,$$

$$s_i(\mu) = \frac{g_{t(n)+m+i}(\mu,t_\sigma)}{g_0(\mu,t_\sigma)} \in \mathbb{R}, \qquad i = 1\ldots,t(n).$$

- PBAS computes X_1-pseudo critical points on $\mathrm{Zer}(\tilde{Q},\mathbb{R}^{m+2t(n)+1})$.

PURDUE UNIVERSITY

Complexity of PBAS

The polynomials $f, g_0, g_1, \ldots, g_{m+2t(n)}$ are

- of degree $O(1)^{m+2t(n)}$ in T;
- of degree $2^{O(m+n^2)}$ in μ.

As a consequence:

Theorem (Basu and M, 2021)

The degree of Zariski closure of the image $\phi((0, \bar{\mu}])$ is $2^{O(m+n^2)}$.

- There exists a bound n^m for generic SDO (Hosten et. al, 2021).

Limit point of CP

The CP system can be viewed as a μ-infinitesimal deformation of

$$\{\mathbf{A}x - b, \quad \mathbf{A}^T y + s - c, \quad (x \otimes_s I_n)s\}.$$

$\mathcal{G} \subset \mathbb{R}[\mu][V]$: the μ-infinitesimally deformed polynomial system.

The set of all μ s.t. a given (f, g) describes $\phi(\mu)$ is semi-algebraic.

- $\exists\ (\check{f}, \check{g})$ and $\check{\mu} > 0$ such that $(\check{f}, \check{g})$ describes $\phi((0, \check{\mu}))$.
- $(x^{**}; y^{**}; s^{**}) \in \lim_\mu(\mathrm{Zer}_b(\mathcal{G}, \mathbb{R}\langle\mu\rangle^{m+2t(n)}))$.

$\mathrm{Zer}_b(\mathcal{G}, \mathbb{R}\langle\mu\rangle^{m+2t(n)})$: bounded zeros of $\mathcal{G}$ over $\mathbb{R}\langle\mu\rangle$.

Our goal is to describe $\lim_\mu(\mathrm{Zer}_b(\mathcal{G}, \mathbb{R}\langle\mu\rangle^{m+2t(n)}))$.

Limit point of CP

We describe $\lim_\mu$ of bounded points represented by $((f,g),\sigma)$:

$$\bar{u} = (\bar{f}, (\bar{g}_0, \bar{g}_1, \ldots, \bar{g}_{m+2t(n)})) \in \mathbb{Z}[T]^{m+2t(n)+2},$$

where

$$\bar{f}(T) = \lim_\mu \mu^{-o(f(T))} f(T),$$

$$\bar{g}_i(T) = \lim_\mu \mu^{-o(g_i^{(k_{\bar{\sigma}}-1)}(T))} g_i^{(k_{\bar{\sigma}}-1)}(T), \quad i = 0, \ldots, m+2t(n).$$

$k_{\bar{\sigma}}$: multiplicity of a real root $t_{\bar{\sigma}}$ of $\bar{f}(T)$ with Thom encoding $\bar{\sigma}$.

$(\bar{u}, \bar{\sigma})$ represents the limit of CP if the ball centered at

$$\bar{x}_i = \frac{\bar{g}_i(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}, \quad \bar{y}_i = \frac{\bar{g}_{t(n)+i}(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}, \quad \bar{s}_i = \frac{\bar{g}_{t(n)+m+i}(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}$$

contains $\phi(\mu)$ in its infinitesimal radius $0 < \delta \ll \mu < 1$.

In summary:

Theorem (Basu and M, 2021)

There exists an algorithm which represents (x^{**}, y^{**}, s^{**}) *as follows*

$$x_i^{**} = \frac{\bar{g}_i(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}, \qquad i = 1\ldots, t(n),$$

$$y_i^{**} = \frac{\bar{g}_{t(n)+i}(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}, \qquad i = 1\ldots, m,$$

$$s_i^{**} = \frac{\bar{g}_{t(n)+m+i}(t_{\bar{\sigma}})}{\bar{g}_0(t_{\bar{\sigma}})}, \qquad i = 1\ldots, t(n),$$

where $t_{\bar{\sigma}}$ is a real root of $\bar{f}(T) = 0$ and $\bar{g}_0(t_{\bar{\sigma}}) \neq 0$.

Now, we have a semi-algebraic description of the convergence rate.

- The quantities $\|X(\mu) - X^{**}\|$ and $\|S(\mu) - S^{**}\|$.

Let $P_x, P_s \in \mathbb{Z}[\mu, D, T_1, T_2]$ defined by

$$P_x = D(g_0(\mu, T_1)\bar{g}_0(T_2))^2 - \sum_{i=1}^{t(n)} (g_i(\mu, T_1)\bar{g}_0(T_2) - \bar{g}_i(T_2)g_0(\mu, T_1))^2,$$

$$P_s = D(g_0(\mu, T_1)\bar{g}_0(T_2))^2 - \sum_{i=m+t(n)+1}^{m+2t(n)} (g_i(\mu, T_1)\bar{g}_0(T_2) - \bar{g}_i(T_2)g_0(\mu, T_1))^2.$$

- $d_x = \|X(\mu) - X^{**}\|^2$ is a root of $P_x(\mu, D, t_\sigma, t_{\bar{\sigma}})$.
- $d_s = \|S(\mu) - S^{**}\|^2$ is a root of $P_s(\mu, D, t_\sigma, t_{\bar{\sigma}})$.

Distance in terms of first-order formulas

In summary, $(\mu, d_x) \in \mathrm{Reali}(\Psi_x, \mathbb{R}^2)$ and $(\mu, d_s) \in \mathrm{Reali}(\Psi_s, \mathbb{R}^2)$:

$$\begin{aligned}\Psi_x(\mu, D) : (\exists T_1)\ (\exists T_2)\ &(P_x(\mu, D, T_1, T_2) = 0) \\ &\wedge\ (f^{(k)}(\mu, T_1) = \sigma(f^{(k)}),\ k = 0, 1, \ldots) \\ &\wedge\ (\bar{f}^{(k)}(T_2) = \bar{\sigma}(\bar{f}^{(k)}),\ k = 0, 1, \ldots),\end{aligned}$$

$$\begin{aligned}\Psi_s(\mu, D) : (\exists T_1)\ (\exists T_2)\ &(P_s(\mu, D, T_1, T_2) = 0) \\ &\wedge\ (f^{(k)}(\mu, T_1) = \sigma(f^{(k)}),\ k = 0, 1, \ldots) \\ &\wedge\ (\bar{f}^{(k)}(T_2) = \bar{\sigma}(\bar{f}^{(k)}),\ k = 0, 1, \ldots).\end{aligned}$$

Using the quantifier elimination algorithm:

Theorem (Basu and M, 2021)

$\|X(\mu) - X^{**}\|, \|S(\mu) - S^{**}\| = O(\mu^{\frac{1}{\gamma}})$, *with* $\gamma = 2^{O(m+n^2)}$.

The idea of the proof

Proof.

- $\mathbb{R}$-realization of the formulas contains a unique (μ, d), $\forall \mu \in (0, \check{\mu})$.

$\Downarrow$

- A q.f. formula involves an atom $R = 0$ with $\deg(R) = 2^{O(m+n^2)}$.

$\Downarrow$

- By the Newton-Puiseux theorem, $R(\mu, D) = 0$ has a root $d \in \mathbb{C}\langle\mu\rangle$.

$\Downarrow$

- The valuation of d is positive and bounded below by $1/\deg_D(R)$.

∎

Summary

We investigated the worst-case convergence of CP:

- Real univariate representations describing CP and its limit point.
  - The complexity of PBAS plays a central rule here.
  - The degree of polynomials leads to a bound $2^{O(m+n^2)}$ on the degree of CP.
- We described the limit point of CP by taking the limit of points in $\mathrm{Zer}_b(\mathcal{G}, \mathbb{R}\langle\mu\rangle^{m+2t(n)})$.

Using quantifier elimination and Newton-Puiseux theorem:

- We provided a bound $O(\mu^{1/\gamma})$, with $\gamma = 2^{O(m+n^2)}$, on the worst-case convergence rate.

Reference: https://arxiv.org/abs/2105.06630

Thank you for your attention

Any questions?