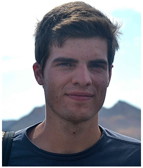
# Hidden Convexity and Benign Non-Convex Landscapes

## ICCOPT 2025

Chenyang Yuan (Joint with Benoît Legat)

Toyota Research Institute

Wednesday 30th July, 2025

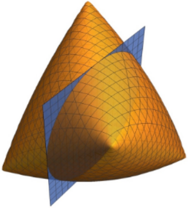


# Introduction

Semidefinite programming (SDP) is a powerful and expressive convex optimization method

Positive semidefinite variable $X \succeq 0$ + linear constraints

Many problems admit SDP relaxations by relaxing low-rank contraints: $X = UU^\top$ to $X \succeq 0$.



**This talk: connections between two areas of low-rank SDP research**

Hidden Convexity:

- Study image of low-rank manifold under structured linear maps
- When are relaxations of low-rank problems exact?

Benign landscape of Burer-Monteiro method:

- Solve SDPs fast using non-convex low-rank formulation
- Will these first-order methods get stuck in local minima?

# Using SDPs for Low-Rank Relaxations

Let $S_+^k$ be the space of $k \times k$ PSD matrices.

We define the following linear map $A : S^k \to \mathbb{R}^m$:

$$A(X) = (\langle A_1, X \rangle, \ldots, \langle A_m, X \rangle)$$

$A(S_+^k)$ is a convex set representable by a SDP

Deciding membership $b \in A(S_+^k)$ is equivalent to feasibility of SDP

$$b = A(X), X \succeq 0$$

Let $S_r^k$ be the space of rank-$r$ PSD matrices

$A(S_r^k)$ is non-convex in general, $A(S_+^k)$ is its convex hull

# Hidden Convexity and Exactness of Relaxations

We say that $A$ has rank-$r$ hidden convexity if following set is convex

$$A(S_+^k) = \{A(X) \mid X \in S_+^k\} = \{A(UU^\top) \mid U \in \mathbb{R}^{k\times r}\} = A(S_r^k)$$

Example: Hidden convexity of the map $A(X) = (\langle C, X\rangle, \langle I, X\rangle)$: $A(S_+^k) = A(S_1^k)$

$$\begin{aligned} \max\ & \langle C, X\rangle \\ \text{s.t.}\ & \mathrm{Tr}(X) = 1 \\ & X \succeq 0 \end{aligned} \quad \text{is equivalent to} \quad \begin{aligned} \max\ & x^\top C x \\ \text{s.t.}\ & \|x\|^2 = 1 \\ & x \in \mathbb{R}^k \end{aligned}$$

SDP relaxation for computing top singular vector is exact

# Examples of Hidden Convexity: Barvinok-Pataki bound

Dines' Theorem: When $A(X) = (\langle C_1, X \rangle, \langle C_2, X \rangle)$ (for any symmetric $C_1, C_2$), $A$ has rank-1 hidden convexity:

$$\left\{(x^\top C_1 x, x^\top C_2 x) \mid x \in \mathbb{R}^k\right\} = \{(\langle C_1, X \rangle, \langle C_2, X \rangle) \mid X \succeq 0\}$$

This can be generalized with the Barvinok-Pataki bound:

When $A(X) = (\langle C_1, X \rangle, \ldots, \langle C_m, X \rangle)$, then $A$ has rank-$r$ hidden convexity when:

$$r \geq \left\lfloor (\sqrt{8m+1} - 1)/2 \right\rfloor$$

# Examples of Hidden Convexity: Sparse SDPs

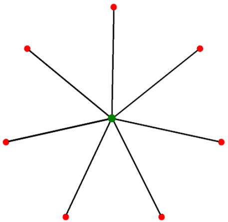
Suppose $C_i$ have an *aggregated sparsity pattern* in terms of graph $G = (V, E)$

$A(X) = (\langle C_1, X\rangle, \ldots, \langle C_m, X\rangle)$ has rank-$\alpha(G)$ hidden convexity [Laurent, Varvitsiotis 2014]

$$\alpha(G) \leq \mathrm{treewidth}(G) + 1$$

Example: When sparsity pattern is a tree, $A(X)$ has rank-2 hidden convexity

$$\begin{bmatrix} * & * & \cdots & * \\ * & * & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ * & 0 & \cdots & * \end{bmatrix}$$



# Examples of Hidden Convexity: Sum of Squares Polynomials

Let $b(x) \in \mathbb{R}^n$ be a basis of $\mathbb{R}[x]_{n,d}$: polynomials in $n$ variables of degree $\leq d$.

$A(X)$ is a map extracting the coefficients of the polynomial $b(x)^\top X b(x)$

Example for $\mathbb{R}[x]_{1,2}$: $b(x) = [1, x, x^2]$,

$$b(x)^\top X b(x) = \left\langle \begin{bmatrix} 1 & x & x^2 \\ x & x^2 & x^3 \\ x^2 & x^3 & x^4 \end{bmatrix}, \begin{bmatrix} X_{11} & X_{12} & X_{13} \\ X_{12} & X_{22} & X_{23} \\ X_{13} & X_{23} & X_{33} \end{bmatrix} \right\rangle$$

$$A(X) = \left( X_{11}, \frac{1}{2} X_{12}, \frac{1}{3}(X_{13} + X_{22}), \frac{1}{2} X_{23}, X_{33} \right)$$

The following are equivalent:

$$p(x) \in \Sigma[x]_{n,2d} \quad \Leftrightarrow \quad p(x) = b(x)^\top X b(x), X \succeq 0 \quad \Leftrightarrow \quad \text{coeff}(p) \in A(S_+^k)$$

Rank of $X$ = the number of squares in sum of squares decomposition

# Pythagoras Numbers

Pythagoras number $\pi(n, d)$: Smallest $r$ so that all polynomials in $\Sigma[x]_{n,2d}$ can be written as sum of $r$ squares in $\mathbb{R}[x]_{n,d}$.

| | $n = 1$ | $n = 2$ | $n = 3$ | $n = 4$ |
|---|---|---|---|---|
| $d = 1$ | 2 [2, 2] | 3 [3, 3] | 4 [4, 4] | 5 [5, 5] |
| $d = 2$ | 2 [2, 2] | 3 [3, 5] | 5 [5, 7] | ? [6, 11] |
| $d = 3$ | 2 [2, 3] | 4 [4, 7] | ? [5, 12] | ? [7, 20] |
| $d = 4$ | 2 [2, 3] | 5 [4, 9] | ? [5, 17] | ? [8, 30] |

Table: Known values of $\pi(n, d)$ [general lower and upper bounds]

- Univariate polynomials $\pi(1, d) = 2$, quadratic forms $\pi(n, 1) = n + 1$
- Ternary quartics $\pi(2, 2) = 3$ [Hilbert 1888]
- $\pi(2, 3) = 4$, $\pi(3, 2) = 5$ [Scheiderer 2017]
- $\pi(2, d) = d + 1$ for $d = 4, 5, 6$ [Blekherman, Dunbar, Sinn 2024]

$A(X)$ for sum of squares decomposition of $\Sigma[x]_{n,2d}$ has rank-$\pi(n, d)$ hidden convexity

# Burer-Monteiro Factorization and Optimization Landscape

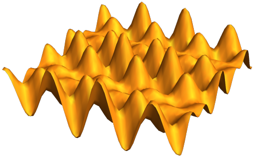
Given $A : S^n \to \mathbb{R}^m$ and $b \in \mathbb{R}^m$, find $X \in S^n_r$ so that $A(X) = b$.

We enforce PSD constraint via factorization $X = UU^\top$

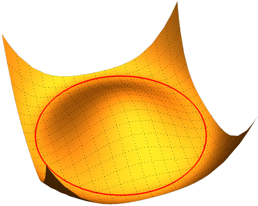
$$\min_{U \in \mathbb{R}^{n \times r}} f_b(U) := \left\| A(UU^\top) - b \right\|^2 .$$

This problem is non-convex in terms of $U$, to be solved with first-order optimization algorithms.

**Is the landscape benign or can we get stuck in local minima?**



**OR**

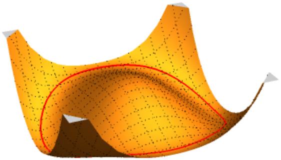


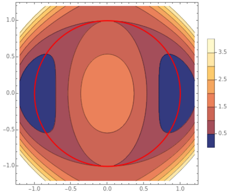
# Optimality Conditions

If $U$ is a second-order critical point of $f_b(U)$, the following conditions must hold for all $V \in \mathbb{R}^{n\times r}$:

$$\langle A(UV^\top), A(UU^\top) - b\rangle \sim \nabla f_b(U)(V) = 0$$
$$\langle A(VV^\top), A(UU^\top) - b\rangle + 2\left\|A(UV^\top)\right\|^2 \sim \nabla^2 f_b(U)(V,V) \geq 0$$

First-order critical points $\supset$ Second-order critical points $\supset \cdots \supset$ Local minima






First order methods (i.e. gradient descent) converge to second-order critical points

If second-order critical point $\implies$ global minima, then gradient descent converges to global minimum, landscape benign

# Prior Work and Our Contribution

Benign landscape for general SDPs: $r \gtrsim \sqrt{m}$ with generic constraints [Cifuentes & Moitra 2019] or smoothed analysis [Bhojanapalli et. al 2018]

Wide body of work for structured SDPs for matrix sensing, synchronization, phase retrieval problems (see `https://sunju.org/research/nonconvex`)

[Zhang 2021] showed that searching for spurious local minima, for fixed spurious point and fixed ground truth in matrix sensing problems, can be formulated as a SDP

Rest of the talk:

- Show that hidden convexity enables search over counterexamples for only fixed spurious point
- Find counterexamples (lower-bounds on $r$) by systematically searching over spurious points
- Analyzing dual formulation leading to proof strategies

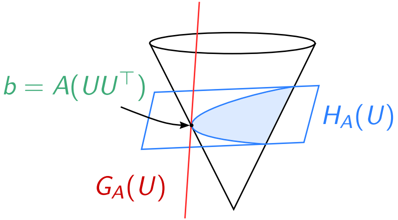
# Certifying Optimality Conditions with SDPs

**Proposition**

*If $A$ has hidden-convexity of rank-$r$, then for fixed $U$, we can certify using a SDP that $f_b(U) = \left\|A(UU^\top) - b\right\|^2$ has no spurious SOCP for all $b \in A(S_r^k)$*

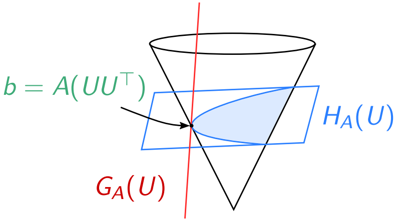
Suffice to show that the following sets only intersect at $b = A(UU^\top)$

$$G_A(U) := \left\{b \in \mathbb{R}^m \mid \nabla f_b(U)(V) = 0, \forall\, V \in \mathbb{R}^{k\times r}\right\}$$
$$H_A(U) := \left\{b \in \mathbb{R}^m \mid \nabla^2 f_b(U)(V,V) \geq 0, \forall\, V \in \mathbb{R}^{k\times r}\right\}$$
$$A(S_r^k) := \left\{A(VV^\top) \in \mathbb{R}^m \mid \forall\, V \in \mathbb{R}^{k\times r}\right\}$$




These are all *convex sets* when $A$ has rank-$r$ hidden convexity!

# Certifying Optimality Conditions with SDPs




$H_A^1 : \mathbb{R}^m \to \mathbb{R}^{kr\times kr}$ a linear map:

$$\mathrm{vec}(V)^\top H_A^1(a)\,\mathrm{vec}(V) := \langle A(VV^\top), a\rangle$$

$H_{A,U}^2 \in \mathbb{R}^{kr\times kr}$ a quadratic form:

$$\mathrm{vec}(V)^\top H_{A,U}^2\,\mathrm{vec}(V) := 2\,\|A_U(V)\|^2$$

We pick random direction $c$ and check if the following SDP has optimum $> 0$

$$\begin{aligned} \max_b\ & \langle c, A(UU^\top) - b\rangle \\ \text{s.t. }\ & A_U^*(A(UU^\top) - b) = 0 \\ & H_A^1(A(UU^\top) - b) + H_{A,U}^2 \succeq 0 \\ & b \in A(S_r^k) \end{aligned} \quad \Leftrightarrow \quad \begin{aligned} \exists b,\gamma \text{ s.t. }\ & \langle c, b\rangle = 1 \\ & A_U^*(\gamma A(UU^\top) - b) = 0 \\ & H_A^1(\gamma A(UU^\top) - b) + \gamma H_{A,U}^2 \succeq 0 \\ & b \in A(S_r^k) \end{aligned}$$

Homogenized version allows us to certify almost surely

# Specializing to Sum of Squares

For polynomials, $\mathcal{A}_{\mathbf{u}} : \mathbb{R}[x]^r_{n,d} \to \mathbb{R}[x]_{n,2d}$

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v}) := \mathcal{A}_{(u_1,\ldots,u_n)}((v_1,\ldots,v_n)) = \sum_i u_i v_i$$

$$\sigma(\mathbf{u}) := \mathcal{A}_{\mathbf{u}}(\mathbf{u})$$

The objective, first- and second-order optimality conditions:

$$f_p(\mathbf{u}) = \|\sigma(\mathbf{u}) - p\|^2$$

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v},\mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle + 2\|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2 \geq 0$$

Prior work:

- Benign landscape for univariate polynomials [Legat, Y., Parrilo 2023]
- Generalization to varieties of minimal degree [Blekherman, Sinn, Velasco, Zhang 2024]

# Dual Certificates

Stronger formulation for projection onto $\Sigma[x]_{n,2d}$: Show that for all $q \in \Sigma[x]_{n,2d}$, following SDP has objective 0

$$\begin{aligned} \max_{p} \; & \langle \sigma(\mathbf{u}) - q, \sigma(\mathbf{u}) - p \rangle \\ \text{s.t. } \forall \mathbf{v} : & \langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle = 0 \\ & \langle \sigma(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle + 2 \left\| \mathcal{A}_{\mathbf{u}}(\mathbf{v}) \right\|^2 \geq 0 \end{aligned}$$

Taking the dual:

$$\min_{\boldsymbol{\lambda}, \mathbf{v_i} \in \mathbb{R}[x]_{n,d}} \sum_{i=1}^{k} \left\| \mathcal{A}_{\mathbf{u}}(\mathbf{v_i}) \right\|^2 \quad \text{s.t.} \quad q - \sigma(\mathbf{u}) = \mathcal{A}_{\mathbf{u}}(\boldsymbol{\lambda}) + \sum_{i=1}^{k} \sigma(\mathbf{v_i})$$

Proof strategy: exhibit a certificate $\boldsymbol{\lambda}, \mathbf{v_i}$ for every $q \in \Sigma[x]_{n,2d}$, $\mathbf{u} \in \mathbb{R}[x]_{n,d}$

**Strong duality may not hold!**

Proof for univariate polynomials constructs an *extended dual certificate* [Legat, Y., Parrilo 2023]

# Automatic Search for Spurious Second-Order Points

Strategy: Given basis of $\mathbb{R}[x]_{n,d}$, search for $\mathbf{u}$ over all subsets of size $r$

Finding the following spurious points $\mathbf{u}$ (where $p = 2x_1^{2d} + \sigma(\mathbf{u})$):

- For $n = 2, d = 2, r = 3$, $\mathbf{u} = (1, x_2, x_2^2)$
- For $n = 2, d = 3, r = 4$, $\mathbf{u} = (1, x_2, x_2^2, x_2^3)$
- For $n = 3, d = 2, r = 5$, $\mathbf{u} = (1, x_2, x_2^2, x_3, x_3^2)$

Generalize examples to following result:

**Proposition**

*For sum of squares decomposition of $\Sigma[x]_{n,2d}$, there are* **spurious second-order critical points** *for any $r \leq \dim(\mathbb{R}[x]_{n-1,d})$*

$n = 2, d = 2$ case appeared in [Blekherman, Sinn, Velasco, Zhang 2024]

# Proof of spurious second-order points

Let $\mathbf{u}$ be any orthogonal basis of $\mathbb{R}[x]_{n-1,d}$ and $\sigma(\mathbf{u}) - p = -2x_1^{2d}$

Since $\mathbf{u}$ does not contain $x_1$, in any monomial of $\mathcal{A}_{\mathbf{u}}(\mathbf{v})$, highest degree of $x_1$ is $d$

First-order condition holds: $\mathcal{A}_{\mathbf{u}}(\mathbf{v})$ has no $x_1^{2d}$ terms, $\langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle = 0$

Write $\mathbf{v} = (c_i x_1^d + \phi_i(x))_{i=1,\ldots,r}$, where $c_i \in \mathbb{R}$ and $\phi_i(x)$ do not have $x_1^d$ terms

$$\langle \sigma(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle = -2 \sum_{i=1}^{r} c_i^2$$

$\mathcal{A}_{\mathbf{u}}(\mathbf{v})$ has only $r$ distinct monomials containing $x_1^d$, each with coefficient $c_i$:

$$2 \left\| \mathcal{A}_{\mathbf{u}}(\mathbf{v}) \right\|^2 \geq 2 \sum_{i=1}^{r} c_i^2$$

Second-order condition holds: $\langle \sigma(\mathbf{v}), \sigma(\mathbf{u}) - p \rangle + 2 \left\| \mathcal{A}_{\mathbf{u}}(\mathbf{v}) \right\|^2 \geq 0$

# Relation to Pythagoras Numbers

Lower bounds on $r$ for benign non-convexity:

| | $n=1$ | $\geq$ | $n=2$ | $\geq$ | $n=3$ | $\geq$ | $n=4$ | $\geq$ |
|---|---|---|---|---|---|---|---|---|
| $d=1$ | 2 [2, 2] | 2 | 3 [3, 3] | 3 | 4 [4, 4] | 4 | 5 [5, 5] | 5 |
| $d=2$ | 2 [2, 2] | 2 | 3 [3, 5] | 4 | 5 [5, 7] | 7 | ? [6, 11] | 11 |
| $d=3$ | 2 [2, 3] | 2 | 4 [4, 7] | 5 | ? [5, 12] | 11 | ? [7, 20] | 21 |
| $d=4$ | 2 [2, 3] | 2 | 5 [4, 9] | 6 | ? [5, 17] | 16 | ? [8, 30] | 36 |
| $d=5$ | 2 [2, 4] | 2 | 6 [4, 11] | 7 | ? [6, 23] | 22 | ? [9, 44] | 57 |
| $d=6$ | 2 [2, 4] | 2 | 7 [4, 13] | 8 | ? [6, 29] | 29 | ? [9, 59] | 85 |

Only true for univariate polynomials and quadratic forms

Provably higher than $\pi(n, d)$ for most other cases!

# Discussion and Conclusion

We showed that hidden convexity enables us to find SDP certificates of no spurious second-order critical point for *fixed* $U$

Computational search to find counterexamples of hidden convexity not implying benign Burer-Monteiro landscape

**Question:** Is there a geometric condition on the map $A(X)$ that leads to both hidden convexity and benign non-convex landscape?

**Question:** Can we automatically find proofs of benign optimization landscapes?