Low-Rank Nonconvex Solver for Sum-of-Squares

SIAMOPT 2023

Benoît Legat (Joint with Chenyang Yuan and Pablo Parrilo)

KU Leuven – MIT

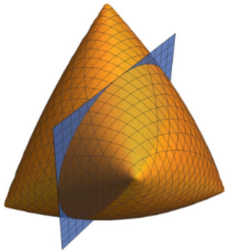
Thursday 1^{st} June, 2023

Introduction

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

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

Solved in polynomial time with interior point methods ($n \sim 10^3$)

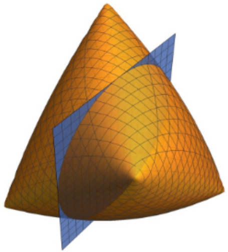


Introduction

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

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

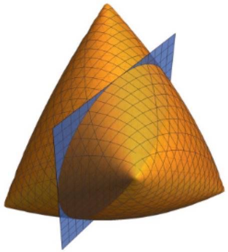
Solved in polynomial time with interior point methods ($n \sim 10^3$)



However: Success of deep learning shows that

- Certain non-convex problems can be solved efficiently in practice with first-order methods ($n > 10^8$)
- Algorithms that scale linearly necessary for working with "big data"

Introduction



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

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

Solved in polynomial time with interior point methods ($n \sim 10^3$)

However: Success of deep learning shows that

- Certain non-convex problems can be solved efficiently in practice with first-order methods ($n > 10^8$)
- Algorithms that scale linearly necessary for working with "big data"

Can we apply these ideas to solving SDPs?

The Burer-Monteiro Method

Burer-Monteiro methods for solving SDPs factor PSD variable $X = UU^T$, then perform local optimization on *non-convex* unconstrained problem

$$\begin{aligned} \langle A_i, X \rangle &= b_i \quad \forall i \\ X &\succeq 0 \end{aligned} \quad \longrightarrow \quad \min_U \sum_i (\langle A_i, UU^T \rangle - b_i)^2$$

$$\text{Feasible} \quad \Longleftrightarrow \quad \text{Optimum} = 0$$

The Burer-Monteiro Method

Burer-Monteiro methods for solving SDPs factor PSD variable $X = UU^T$, then perform local optimization on *non-convex* unconstrained problem

$$\begin{aligned} \langle A_i, X \rangle &= b_i \quad \forall i \\ X &\succeq 0 \end{aligned} \quad \longrightarrow \quad \min_U \sum_i (\langle A_i, UU^T \rangle - b_i)^2$$

$$\text{Feasible} \quad \Longleftrightarrow \quad \text{Optimum} = 0$$

May get stuck in local optimum (explicit counterexamples where second-order critical point $\neq$ global minimum)

The Burer-Monteiro Method

Burer-Monteiro methods for solving SDPs factor PSD variable $X = UU^T$, then perform local optimization on *non-convex* unconstrained problem

$$\begin{aligned} \langle A_i, X \rangle &= b_i \quad \forall i \\ X &\succeq 0 \end{aligned} \quad \longrightarrow \quad \min_U \sum_i (\langle A_i, UU^T \rangle - b_i)^2$$

$$\text{Feasible} \quad \Longleftrightarrow \quad \text{Optimum} = 0$$

May get stuck in local optimum (explicit counterexamples where second-order critical point $\neq$ global minimum)

OR

When is non-convexity benign?

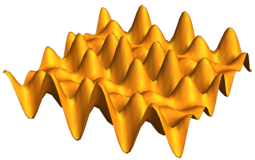
The Burer-Monteiro Method

Burer-Monteiro methods for solving SDPs factor PSD variable $X = UU^T$, then perform local optimization on *non-convex* unconstrained problem

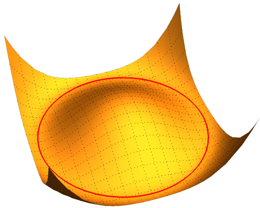
$$\begin{aligned} \langle A_i, X \rangle &= b_i \quad \forall i \\ X &\succeq 0 \end{aligned} \quad \longrightarrow \quad \min_U \sum_i (\langle A_i, UU^T \rangle - b_i)^2$$

$$\text{Feasible} \iff \text{Optimum} = 0$$

May get stuck in local optimum (explicit counterexamples where second-order critical point $\neq$ global minimum)



OR



When is non-convexity benign?

Related Work

SDP with m linear constraints, factorization $X = UU^\top$, where $U \in \mathbb{R}^{n \times r}$.

[BM05] Burer and Monteiro. "Local Minima and Convergence in Low-Rank Semidefinite Programming". 2005.

[CM19] Cifuentes and Moitra. "Polynomial Time Guarantees for the Burer-Monteiro Method". 2019.

[Bho+18] Bhojanapalli et al. "Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form". 2018.

[GJZ17] Ge, Jin, and Zheng. "No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis". 2017.

[BBV16] Bandeira, Boumal, and Voroninski. "On the low-rank approach for semidefinite programs arising in synchronization and community detection". 2016.

Related Work

SDP with m linear constraints, factorization $X = UU^\top$, where $U \in \mathbb{R}^{n \times r}$.

Second-order critical points $\implies$ **Global minima** (non-convexity benign):

- $r \geq n$ [BM05] (explicit counterexamples exist for $r = n - 1$, $m = n$)
- $r \gtrsim \sqrt{m}$ with smoothed analysis [CM19], determinant regularization [BM05] or generic constraints [Bho+18]
- $r \gtrsim r^*$, where r^* maximum possible rank of SDP solution (matrix sensing [GJZ17], rotational synchronization [BBV16])

[BM05] Burer and Monteiro. "Local Minima and Convergence in Low-Rank Semidefinite Programming". 2005.

[CM19] Cifuentes and Moitra. "Polynomial Time Guarantees for the Burer-Monteiro Method". 2019.

[Bho+18] Bhojanapalli et al. "Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form". 2018.

[GJZ17] Ge, Jin, and Zheng. "No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis". 2017.

[BBV16] Bandeira, Boumal, and Voroninski. "On the low-rank approach for semidefinite programs arising in synchronization and community detection". 2016.

Related Work

SDP with m linear constraints, factorization $X = UU^\top$, where $U \in \mathbb{R}^{n \times r}$.

Second-order critical points $\implies$ **Global minima** (non-convexity benign):

- $r \geq n$ [BM05] (explicit counterexamples exist for $r = n - 1$, $m = n$)
- $r \gtrsim \sqrt{m}$ with smoothed analysis [CM19], determinant regularization [BM05] or generic constraints [Bho+18]
- $r \gtrsim r^*$, where r^* maximum possible rank of SDP solution (matrix sensing [GJZ17], rotational synchronization [BBV16])

Smaller r in factorization $\to$ less benign landscape

[BM05] Burer and Monteiro. "Local Minima and Convergence in Low-Rank Semidefinite Programming". 2005.

[CM19] Cifuentes and Moitra. "Polynomial Time Guarantees for the Burer-Monteiro Method". 2019.

[Bho+18] Bhojanapalli et al. "Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form". 2018.

[GJZ17] Ge, Jin, and Zheng. "No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis". 2017.

[BBV16] Bandeira, Boumal, and Voroninski. "On the low-rank approach for semidefinite programs arising in synchronization and community detection". 2016.

Related Work

SDP with m linear constraints, factorization $X = UU^\top$, where $U \in \mathbb{R}^{n \times r}$.

Second-order critical points $\implies$ **Global minima** (non-convexity benign):

- $r \geq n$ [BM05] (explicit counterexamples exist for $r = n - 1$, $m = n$)
- $r \gtrsim \sqrt{m}$ with smoothed analysis [CM19], determinant regularization [BM05] or generic constraints [Bho+18]
- $r \gtrsim r^*$, where r^* maximum possible rank of SDP solution (matrix sensing [GJZ17], rotational synchronization [BBV16])

Smaller r in factorization $\to$ less benign landscape

Can we get do better if the SDP has special structure?

[BM05] Burer and Monteiro. "Local Minima and Convergence in Low-Rank Semidefinite Programming". 2005.

[CM19] Cifuentes and Moitra. "Polynomial Time Guarantees for the Burer-Monteiro Method". 2019.

[Bho+18] Bhojanapalli et al. "Smoothed analysis for low-rank solutions to semidefinite programs in quadratic penalty form". 2018.

[GJZ17] Ge, Jin, and Zheng. "No Spurious Local Minima in Nonconvex Low Rank Problems: A Unified Geometric Analysis". 2017.

[BBV16] Bandeira, Boumal, and Voroninski. "On the low-rank approach for semidefinite programs arising in synchronization and community detection". 2016.

Sum of Squares Optimization

Given $p(x)$, can we write it as a sum of squares? $p(x) = \sum_{i=1}^{r} u_i(x)^2$

Certifies that $p(x) \geq 0$, and can be formulated as a SDP:

$$p(x) = \vec{b}(x)^\top Q \vec{b}(x), \quad Q \succeq 0$$

[Chu+16] Chua et al. "Gram spectrahedra". 2016.

Sum of Squares Optimization

Given $p(x)$, can we write it as a sum of squares? $p(x) = \sum_{i=1}^{r} u_i(x)^2$

Certifies that $p(x) \geq 0$, and can be formulated as a SDP:

$$p(x) = \vec{b}(x)^\top Q \vec{b}(x), \quad Q \succeq 0$$

Q satisfying above constraints is called the Gram spectrahedron [Chu+16]

[Chu+16] Chua et al. "Gram spectrahedra". 2016.

Sum of Squares Optimization

Previous work: rank needed for benign non-convexity $\sim$ max rank of extreme points of Gram spectrahedron

Can we do better?

[Sch22] Scheiderer. "Extreme points of Gram spectrahedra of binary forms". 2022.

Image credit: Tae Roh and Lieven Vandenberghe. (2006) Discrete transforms, semidefinite programming and sum-of-squares representations of nonnegative polynomials. SIAM J. on Optimization.

Sum of Squares Optimization

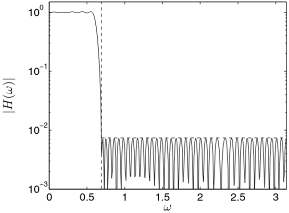
Previous work: rank needed for benign non-convexity $\sim$ max rank of extreme points of Gram spectrahedron

Can we do better?

Univariate (trigonometric) polynomials:

$$p(x) = a_0 + \sum_{k=1}^{2d} a_k \cos(kx) \quad x \in [0, 2\pi]$$

Applications in signal processing, filter design and control



[Sch22] Scheiderer. "Extreme points of Gram spectrahedra of binary forms". 2022.

Image credit: Tae Roh and Lieven Vandenberghe. (2006) Discrete transforms, semidefinite programming and sum-of-squares representations of nonnegative polynomials. SIAM J. on Optimization.

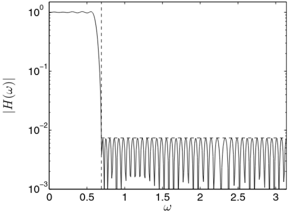
Sum of Squares Optimization

Previous work: rank needed for benign non-convexity $\sim$ max rank of extreme points of Gram spectrahedron

Can we do better?

Univariate (trigonometric) polynomials:

$$p(x) = a_0 + \sum_{k=1}^{2d} a_k \cos(kx) \quad x \in [0, 2\pi]$$



Applications in signal processing, filter design and control

Gram spectrahedra has extreme points of all ranks: $2 \le r \lesssim \sqrt{d}$ [Sch22]

But always has rank-2 point! (Sum of 2 squares)

[Sch22] Scheiderer. "Extreme points of Gram spectrahedra of binary forms". 2022.

Image credit: Tae Roh and Lieven Vandenberghe. (2006) Discrete transforms, semidefinite programming and sum-of-squares representations of nonnegative polynomials. SIAM J. on Optimization.

Contributions

Find sum of squares decomposition of $p(x)$ by solving (equivalent to B-M):

$$\min_{\mathbf{u}} f_p(\mathbf{u}) = \left\| p(x) - \sum_{i=1}^r u_i(x)^2 \right\|^2$$

Contributions

Find sum of squares decomposition of $p(x)$ by solving (equivalent to B-M):

$$\min_{\mathbf{u}} f_p(\mathbf{u}) = \left\| p(x) - \sum_{i=1}^{r} u_i(x)^2 \right\|^2$$

For any norm on polynomials, if $f_p(\mathbf{u}) = 0$, $\sum_i u_i(x)^2$ is a sum of squares decomposition of $p(x)$.

Contributions

Find sum of squares decomposition of $p(x)$ by solving (equivalent to B-M):

$$\min_{\mathbf{u}} f_p(\mathbf{u}) = \left\| p(x) - \sum_{i=1}^{r} u_i(x)^2 \right\|^2$$

For any norm on polynomials, if $f_p(\mathbf{u}) = 0$, $\sum_i u_i(x)^2$ is a sum of squares decomposition of $p(x)$.

Theorem

For all nonnegative univariate polynomials $p(x) \in \mathbb{R}[x]_{2d}$ and any $r \geq 2$, if $\mathbf{u} \in \mathbb{R}[x]_d^r$ satisfies $\nabla f_p(\mathbf{u}) = 0$ and $\nabla^2 f_p(\mathbf{u}) \succeq 0$, then $f_p(\mathbf{u}) = 0$.

Contributions

Find sum of squares decomposition of $p(x)$ by solving (equivalent to B-M):

$$\min_{\mathbf{u}} f_p(\mathbf{u}) = \left\| p(x) - \sum_{i=1}^r u_i(x)^2 \right\|^2$$

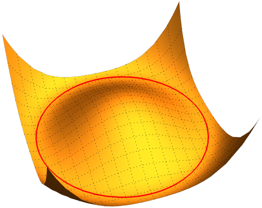
For any norm on polynomials, if $f_p(\mathbf{u}) = 0$, $\sum_i u_i(x)^2$ is a sum of squares decomposition of $p(x)$.

Theorem

For all nonnegative univariate polynomials $p(x) \in \mathbb{R}[x]_{2d}$ and any $r \geq 2$, if $\mathbf{u} \in \mathbb{R}[x]_d^r$ satisfies $\nabla f_p(\mathbf{u}) = 0$ and $\nabla^2 f_p(\mathbf{u}) \succeq 0$, then $f_p(\mathbf{u}) = 0$.

First-order methods find sum of squares decomposition (non-convexity benign)

If we choose a suitable norm, $\nabla f_p(\mathbf{u})$ can be computed in $O(d \log d)$ time using Fast Fourier Transforms (FFTs)



Proof Sketch

Define Sylvester map $\mathcal{A}_{\mathbf{u}} : \mathbb{R}[x]_d^r \to \mathbb{R}[x]_{2d}$

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v}) = \mathcal{A}_{(u_1, u_2)}((v_1, v_2)) = u_1 v_1 + u_2 v_2$$

Proof Sketch

Define Sylvester map $\mathcal{A}_{\mathbf{u}} : \mathbb{R}[x]_d^r \to \mathbb{R}[x]_{2d}$

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v}) = \mathcal{A}_{(u_1,u_2)}((v_1, v_2)) = u_1 v_1 + u_2 v_2$$

Given an inner product $\langle \cdot, \cdot \rangle$ on polynomials with associated norm $\|\cdot\|$:

$$f_p(\mathbf{u}) = \left\| u_1^2 + u_2^2 - p \right\|^2$$
$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle$$
$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2$$

Proof Sketch

Define Sylvester map $\mathcal{A}_{\mathbf{u}} : \mathbb{R}[x]_d^r \to \mathbb{R}[x]_{2d}$

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v}) = \mathcal{A}_{(u_1,u_2)}((v_1, v_2)) = u_1 v_1 + u_2 v_2$$

Given an inner product $\langle \cdot, \cdot \rangle$ on polynomials with associated norm $\|\cdot\|$:

$$f_p(\mathbf{u}) = \left\| u_1^2 + u_2^2 - p \right\|^2$$
$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle$$
$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2$$

Goal: For all $\mathbf{u}$ such that $\nabla f_p(\mathbf{u})(\mathbf{v}) = 0$ and $\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \succeq 0$ for all $\mathbf{v}$, show that $f_p(\mathbf{u}) = 0$.

Proof Sketch

Define Sylvester map $\mathcal{A}_{\mathbf{u}} : \mathbb{R}[x]_d^r \to \mathbb{R}[x]_{2d}$

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v}) = \mathcal{A}_{(u_1,u_2)}((v_1, v_2)) = u_1 v_1 + u_2 v_2$$

Given an inner product $\langle \cdot, \cdot \rangle$ on polynomials with associated norm $\|\cdot\|$:

$$\begin{aligned} f_p(\mathbf{u}) &= \left\| u_1^2 + u_2^2 - p \right\|^2 \\ \nabla f_p(\mathbf{u})(\mathbf{v}) &\sim \left\langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle \\ \nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) &\sim \left\langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2 \end{aligned}$$

Goal: For all $\mathbf{u}$ such that $\nabla f_p(\mathbf{u})(\mathbf{v}) = 0$ and $\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \succeq 0$ for all $\mathbf{v}$, show that $f_p(\mathbf{u}) = 0$.

To do so, for every $p \in \Sigma[x]_{2d}$ and $\mathbf{u} \in \mathbb{R}[x]_d^2$, find $\mathbf{v_i} \in \mathbb{R}[x]_d^2$ so that:

$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = -\left\| u_1^2 + u_2^2 - p \right\|^2 = -f_p(\mathbf{u})$$

Geometric Interpretation

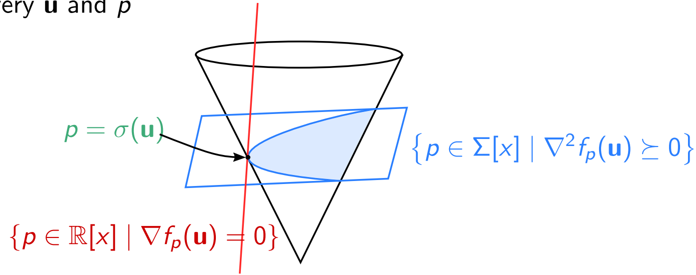
$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = -\left\|u_1^2 + u_2^2 - p\right\|^2 = -f_p(\mathbf{u})$$

Our proof can be interpreted as finding a Positivstellensatz certificate of this condition for every $\mathbf{u}$ and p

Geometric Interpretation

$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = -\left\|u_1^2 + u_2^2 - p\right\|^2 = -f_p(\mathbf{u})$$

Our proof can be interpreted as finding a Positivstellensatz certificate of this condition for every $\mathbf{u}$ and p

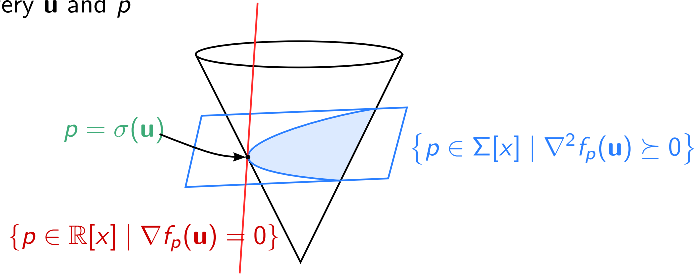


Geometrically, we want to show that the only intersection between sets with zero gradient and PSD hessian is when $f_p(\mathbf{u}) = 0$.

Geometric Interpretation

$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = -\left\| u_1^2 + u_2^2 - p \right\|^2 = -f_p(\mathbf{u})$$

Our proof can be interpreted as finding a Positivstellensatz certificate of this condition for every $\mathbf{u}$ and p



Geometrically, we want to show that the only intersection between sets with zero gradient and PSD hessian is when $f_p(\mathbf{u}) = 0$.

For fixed $\mathbf{u}$, these sets are convex (and can be represented by SDPs)!

Proof Sketch

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2 \geq 0$$

$$-\left\| u_1^2 + u_2^2 - p \right\|^2 = \nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i})$$

Proof Sketch

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2 \geq 0$$

$$-\|u_1^2 + u_2^2 - p\|^2 = \nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i})$$

Suppose u_1, u_2 are coprime (true generically)

Proof Sketch

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \rangle + \|\mathcal{A}_{\mathbf{u}}(\mathbf{v})\|^2 \ge 0$$

$$-\left\| u_1^2 + u_2^2 - p \right\|^2 = \nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i})$$

Suppose u_1, u_2 are coprime (true generically)

Bézout's lemma ($\mathcal{A}_{\mathbf{u}}$ is onto) $\implies$ there exist $\mathbf{v_0}$ such that

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v_0}) = -(u_1^2 + u_2^2 - p) \implies \nabla f_p(\mathbf{u})(\mathbf{v_0}) = -\left\| u_1^2 + u_2^2 - p \right\|^2$$

Proof Sketch

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{u}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \left\langle \mathcal{A}_{\mathbf{v}}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle + \left\| \mathcal{A}_{\mathbf{u}}(\mathbf{v}) \right\|^2 \geq 0$$

$$-\left\| u_1^2 + u_2^2 - p \right\|^2 = \nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i})$$

Suppose u_1, u_2 are coprime (true generically)

Bézout's lemma ($\mathcal{A}_{\mathbf{u}}$ is onto) $\implies$ there exist $\mathbf{v_0}$ such that

$$\mathcal{A}_{\mathbf{u}}(\mathbf{v_0}) = -(u_1^2 + u_2^2 - p) \implies \nabla f_p(\mathbf{u})(\mathbf{v_0}) = -\left\| u_1^2 + u_2^2 - p \right\|^2$$

Suppose $u_1 = u_2$. If $p(x) = 2\sum_i s_i(x)^2$, choose $\mathbf{v_i} = (s_i, -s_i)$, $\mathbf{v_0} = (-u_1, -u_2)$:

$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) = -\left\langle u_1^2 + u_2^2, u_1^2 + u_2^2 - p \right\rangle$$

$$\sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = \left\langle p, u_1^2 + u_2^2 - p \right\rangle$$

Proof Sketch

$$\nabla f_p(\mathbf{u})(\mathbf{v}) \sim \left\langle \mathcal{A}_\mathbf{u}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle = 0$$

$$\nabla^2 f_p(\mathbf{u})(\mathbf{v}, \mathbf{v}) \sim \left\langle \mathcal{A}_\mathbf{v}(\mathbf{v}), u_1^2 + u_2^2 - p \right\rangle + \|\mathcal{A}_\mathbf{u}(\mathbf{v})\|^2 \geq 0$$

$$-\left\| u_1^2 + u_2^2 - p \right\|^2 = \nabla f_p(\mathbf{u})(\mathbf{v_0}) + \sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i})$$

Suppose u_1, u_2 are coprime (true generically)

Bézout's lemma ($\mathcal{A}_\mathbf{u}$ is onto) $\implies$ there exist $\mathbf{v_0}$ such that

$$\mathcal{A}_\mathbf{u}(\mathbf{v_0}) = -(u_1^2 + u_2^2 - p) \implies \nabla f_p(\mathbf{u})(\mathbf{v_0}) = -\left\| u_1^2 + u_2^2 - p \right\|^2$$

Suppose $u_1 = u_2$. If $p(x) = 2\sum_i s_i(x)^2$, choose $\mathbf{v_i} = (s_i, -s_i)$, $\mathbf{v_0} = (-u_1, -u_2)$:

$$\nabla f_p(\mathbf{u})(\mathbf{v_0}) = -\left\langle u_1^2 + u_2^2, u_1^2 + u_2^2 - p \right\rangle$$

$$\sum_{i=1}^{k} \nabla^2 f_p(\mathbf{u})(\mathbf{v_i}, \mathbf{v_i}) = \left\langle p, u_1^2 + u_2^2 - p \right\rangle$$

Main technical result: how to interpolate between these two cases

Sampled Basis

Theorem holds for any inner product $\langle p(x), q(x) \rangle$ on polynomials, which should we choose?

[LP04] Lofberg and Parrilo. "From Coefficients to Samples: A New Approach to SOS Optimization". 2004.
[CP17] Cifuentes and Parrilo. "Sampling Algebraic Varieties for Sum of Squares Programs". 2017.

Sampled Basis

Theorem holds for any inner product $\langle p(x), q(x) \rangle$ on polynomials, which should we choose?

Given $p(x), q(x)$ of degree d, choose $d+1$ points x_k

$$\langle p(x), q(x) \rangle = \sum_{k=1}^{d+1} p(x_k) q(x_k), \quad \|p(x)\|^2 = \sum_{k=1}^{d+1} p(x_k)^2$$

Valid inner product: when x_k distinct, if $\|p(x)\|^2 = 0$ then $p(x) = 0$.

[LP04] Lofberg and Parrilo. "From Coefficients to Samples: A New Approach to SOS Optimization". 2004.

[CP17] Cifuentes and Parrilo. "Sampling Algebraic Varieties for Sum of Squares Programs". 2017.

Sampled Basis

Theorem holds for any inner product $\langle p(x), q(x)\rangle$ on polynomials, which should we choose?

Given $p(x), q(x)$ of degree d, choose $d+1$ points x_k

$$\langle p(x), q(x)\rangle = \sum_{k=1}^{d+1} p(x_k)q(x_k), \quad \|p(x)\|^2 = \sum_{k=1}^{d+1} p(x_k)^2$$

Valid inner product: when x_k distinct, if $\|p(x)\|^2 = 0$ then $p(x) = 0$.

Sum of squares using a sampled/interpolation basis studied by [LP04] and [CP17].

[LP04] Lofberg and Parrilo. "From Coefficients to Samples: A New Approach to SOS Optimization". 2004.

[CP17] Cifuentes and Parrilo. "Sampling Algebraic Varieties for Sum of Squares Programs". 2017.

Sampled Basis

Theorem holds for any inner product $\langle p(x), q(x) \rangle$ on polynomials, which should we choose?

Given $p(x), q(x)$ of degree d, choose $d+1$ points x_k

$$\langle p(x), q(x) \rangle = \sum_{k=1}^{d+1} p(x_k) q(x_k), \quad \|p(x)\|^2 = \sum_{k=1}^{d+1} p(x_k)^2$$

Valid inner product: when x_k distinct, if $\|p(x)\|^2 = 0$ then $p(x) = 0$.

Sum of squares using a sampled/interpolation basis studied by [LP04] and [CP17].

How should we choose x_k?

[LP04] Lofberg and Parrilo. "From Coefficients to Samples: A New Approach to SOS Optimization". 2004.

[CP17] Cifuentes and Parrilo. "Sampling Algebraic Varieties for Sum of Squares Programs". 2017.

Numerical Implementation

Compute sum of squares decomposition of degree $4n$ trigonometric polynomial

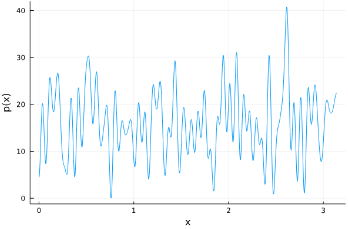
$$p(x) = a_0 + \sum_{k=1}^{2d} a_k \cos(kx) \quad x \in [0, \pi]$$

Using basis vectors evaluated at $4d + 1$ points

$$B_k = [1, \cos(x_k), \ldots, \cos(dx_k)]$$
$$x_k = \frac{k\pi}{d}, \quad k = 1, \ldots, 4d + 1$$

Matrix-vector producted in $\nabla f_p(U)$ computed by FFT

$$\nabla f_p(U) = U^\top B \,\mathrm{Diag}(\left\|U^\top B_k\right\|^2 - p(x_k))B^\top$$



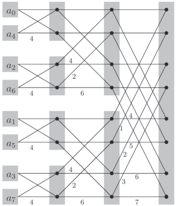
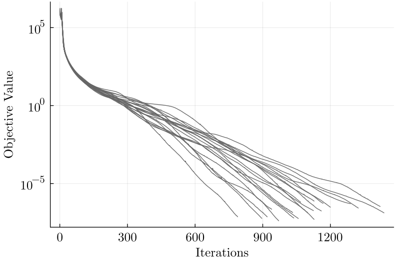


Image credit: Christos Papadimitriou, Sanjoy Dasgupta, and Umesh Vazirani. (2006) Algorithms

Numerical Results

Compute sum of squares decomposition for random trigonometric polynomial

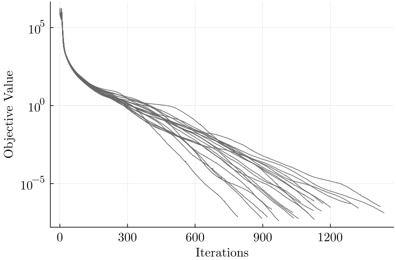
Convergence rate for L-BFGS with random initialization



Numerical Results

Compute sum of squares decomposition for random trigonometric polynomial

Convergence rate for L-BFGS with random initialization



Results (stop at 10^{-7} relative error in $\mathbf{u}$):

Degree of $p(x)$	10,000	20,000	100,000	200,000	1,000,000
Time (s)	6	9	53	160	1461
Iterations	530	632	1126	1375	2303

General Sum-of-Squares

$$\begin{array}{rl} \min_{u_i} & \langle \mu_0, (\sum_{i=1}^r u_i(x)^2) \rangle \\ \text{s.t.} & \langle \mu_i, (\sum_{i=1}^r u_i(x)^2) \rangle = b_i \quad i = 1, \dots m \end{array}$$

Lagrangian

$$\begin{aligned} \nu(y) &= \mu_0 - \sum y_i \mu_i \\ \mathcal{L}(\mathbf{u}, y) &= \langle \mu_0, \sum u_i^2 \rangle - \sum y_i (\langle \mu_i, \sum u_i^2 \rangle - b_i) \\ &= \langle \nu(y), \sum u_i^2 \rangle + \langle b, y \rangle \, . \end{aligned}$$

Dual certificate

If $\mathbf{u}$ feasible, $\langle \nu(y), \sum u_i^2 \rangle = 0$ and $\langle \nu(y), v^2 \rangle \geq 0, \forall v$ then $\forall \mathbf{w}$ feasible,

$$\langle \mu_0, \sum w_i \rangle = \mathcal{L}(\mathbf{w}, y) \geq \langle b, y \rangle = \mathcal{L}(\mathbf{u}, y) = \langle \mu_0, \sum u_i \rangle$$

Manifold

Feasible set as manifold ?

$$\mathcal{M} = \left\{ \mathbf{u} \in \mathbb{R}[x]_d \mid \langle \mu_i, \left(\sum_{i=1}^r u_i(x)^2 \right) \rangle = b_i,\ i = 1, \ldots, m \right\}.$$

Theorem: If $u_i(x) \in \mathcal{M}$ and *manifold*-SOCP then $\langle \nu(y), v^2 \rangle \geq 0\ \forall v$.

Corollary: There is no spurious local minimum.

Can we use Riemanian manifold optimization algorithms ?

Rank of tangent space depends depends on $\gcd(u_1(x), \ldots, u_r(x))$ so $\mathcal{M}$ is not *Riemannian*.

Conclusion

When does it make sense to solve non-convex formulations of convex problems?

In our setting we can prove that non-convexity does not hurt us

Near-linear time iteration cost with first-order methods in a benign landscape