Polynomial optimization on finite sets.

Mauricio Velasco
Universidad de los Andes (Colombia)

Current themes of Discrete Optimization:
Boot-camp for early career researchers
ICERM, January 2023

Polynomial Optimization

Let $X \subseteq \mathbb{R}^n$ be a finite set of points.

Polynomial Optimization

Let $X \subseteq \mathbb{R}^n$ be a finite set of points.

The **polynomial optimization** *problem on X can be stated as: Given a polynomial $f(x_1, \ldots, x_n) \in \mathbb{R}[x_1, \ldots, x_n]$ find the number*

$$f_{\min} := \min_{x \in X} f(x)$$

Many important problems are special cases...

If $X = \{0,1\}^n$, L is a weighted graph with edge weights w_{ij} and

$$f(x_1, \ldots, x_n) := - \sum_{(i,j) \in E(L)} w_{ij}(x_i - x_j)^2$$

Many important problems are special cases...

If $X = \{0,1\}^n$, L is a weighted graph with edge weights w_{ij} and

$$f(x_1, \dots, x_n) := -\sum_{(i,j)\in E(L)} w_{ij}(x_i - x_j)^2$$

then $f_{\min} = -\mathrm{MaxCut}(L)$.

Many important problems are special cases...

If $X = \{0,1\}^n$, L is a weighted graph with edge weights w_{ij} and

$$f(x_1, \dots, x_n) := -\sum_{(i,j)\in E(L)} w_{ij}(x_i - x_j)^2$$

then $f_{\min} = -\mathrm{MaxCut}(L)$.

Optimization over finite subsets of $\mathbb{R}^n$ is hard.

A reframing of the problem...

Let $X \subseteq \mathbb{R}^n$ be a finite set of points and $f(x_1, \ldots, x_n)$ a polynomial with real coefficients. We want

$$f_{\min} := \min_{x \in X} f(x)$$

A reframing of the problem...

Let $X \subseteq \mathbb{R}^n$ be a finite set of points and $f(x_1, \ldots, x_n)$ a polynomial with real coefficients. We want

$$f_{\min} := \min_{x \in X} f(x)$$

Lemma.

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

where P is the collection of polynomials which are nonnegative at all points of X.

A reformulation of the problem...

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

A reformulation of the problem...

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

1. We shift our attention from the points of X to the **functions on the finite set** X. Such functions form a ring $\mathbb{R}[X]$.

A reformulation of the problem...

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

1. We shift our attention from the points of X to the **functions on the finite set** X. Such functions form a ring $\mathbb{R}[X]$.
2. We focus on **the convex cone** P **consisting of functions that are nonnegative at all points of** X. If we had a fast membership algorithm for P then we could solve optimization problems via binary search.

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\textit{Polynomial functions in } \mathbb{R}^n \textit{ restricted to } X\}$$

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\text{Polynomial functions in } \mathbb{R}^n \text{ restricted to } X\}$$

Example:

In $X = \{-1, 1\}^3 \subseteq \mathbb{R}^3$ we restrict polynomials in x, y, z to X.

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\textit{Polynomial functions in } \mathbb{R}^n \textit{ restricted to } X\}$$

Example:

In $X = \{-1, 1\}^3 \subseteq \mathbb{R}^3$ we restrict polynomials in x, y, z to X.
How much is $x^2 = ?$

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\text{Polynomial functions in } \mathbb{R}^n \text{ restricted to } X\}$$

Example:

In $X = \{-1,1\}^3 \subseteq \mathbb{R}^3$ we restrict polynomials in x, y, z to X.
How much is $x^2 = ?$

$$x^2 = 1 \text{ in } \mathbb{R}[X]$$

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ *is*

$$\mathbb{R}[X] := \{\textit{Polynomial functions in } \mathbb{R}^n \textit{ restricted to } X\}$$

Example:

In $X = \{-1, 1\}^3 \subseteq \mathbb{R}^3$ *we restrict polynomials in* x, y, z *to* X.
How much is $x^2 =$*?*

$$x^2 = 1 \textit{ in } \mathbb{R}[X]$$

How about

$$(x - y)^2 + (x - z)^2 + (y - z)^2 =$$

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\textit{Polynomial functions in } \mathbb{R}^n \textit{ restricted to } X\}$$

Example:

In $X = \{-1, 1\}^3 \subseteq \mathbb{R}^3$ we restrict polynomials in x, y, z to X.
How much is $x^2 =$?

$$x^2 = 1 \textit{ in } \mathbb{R}[X]$$

How about

$$(x-y)^2 + (x-z)^2 + (y-z)^2 = x^2 - 2xy + y^2 + \dots$$

1. Polynomial functions on X

Definition.

The coordinate ring of a set $X \subseteq \mathbb{R}^n$ is

$$\mathbb{R}[X] := \{\textit{Polynomial functions in } \mathbb{R}^n \textit{ restricted to } X\}$$

Example:

In $X = \{-1,1\}^3 \subseteq \mathbb{R}^3$ we restrict polynomials in x, y, z to X.
How much is $x^2 =$?

$$x^2 = 1 \textit{ in } \mathbb{R}[X]$$

How about

$$(x-y)^2 + (x-z)^2 + (y-z)^2 = x^2 - 2xy + y^2 + \dots$$

$$= (2-2xy) + (2-2xz) + (2-2yz) = 6 - 2(xy + xz + yz)$$

1. Polynomial functions on X

For every finite set $X \subseteq \mathbb{R}^n$ there is a finite collection of polynomial equations $h_1 = 0, h_2 = 0, \ldots, h_M = 0$ with $h_i \in \mathbb{R}[x_1, \ldots, x_n]$ which generate all relations in the sense that

$$\mathbb{R}[X] = \mathbb{R}[x_1, \ldots, x_n]/(h_1, \ldots, h_m)$$

where

$$(h_1, \ldots, h_M) := \{p_1 h_1 + \cdots + p_M h_M : p_i \in \mathbb{R}[\vec{x}]\}$$

1. Polynomial functions on X

For every finite set $X \subseteq \mathbb{R}^n$ there is a finite collection of polynomial equations $h_1 = 0, h_2 = 0, \ldots, h_M = 0$ with $h_i \in \mathbb{R}[x_1, \ldots, x_n]$ which generate all relations in the sense that

$$\mathbb{R}[X] = \mathbb{R}[x_1, \ldots, x_n]/(h_1, \ldots, h_m)$$

where

$$(h_1, \ldots, h_M) := \{p_1 h_1 + \cdots + p_M h_M : p_i \in \mathbb{R}[\vec{x}]\}$$

Example:

If $X = \{-1, 1\}^3 \subseteq \mathbb{R}^3$ *then*

$$\mathbb{R}[X] = \mathbb{R}[x, y, z]/(x^2 - 1, y^2 - 1, z^2 - 1).$$

1. Polynomial functions on X

Which functions $s : X \to \mathbb{R}$ are the restriction of a polynomial (i.e. live in $\mathbb{R}[X]$)?

1. Polynomial functions on X

Which functions $s : X \to \mathbb{R}$ are the restriction of a polynomial (i.e. live in $\mathbb{R}[X]$)?

Theorem.

If X is finite then $\mathbb{R}[X]$ is precisely the set of all real-valued functions on X. In particular $\dim(\mathbb{R}[X]) = |X|$.

1. Polynomial functions on X

Which functions $s : X \to \mathbb{R}$ are the restriction of a polynomial (i.e. live in $\mathbb{R}[X]$)?

Theorem.

If X is finite then $\mathbb{R}[X]$ is precisely the set of all real-valued functions on X. In particular $\dim(\mathbb{R}[X]) = |X|$.

The **Lagrange interpolation functions** $\{p_z\}_{z \in X}$, defined by

$$p_z(x) := \begin{cases} 1, & \text{if } x = z \text{ and} \\ 0, & \text{if } x \neq z \end{cases}$$

1. Polynomial functions on X

Which functions $s : X \to \mathbb{R}$ are the restriction of a polynomial (i.e. live in $\mathbb{R}[X]$)?

Theorem.

If X is finite then $\mathbb{R}[X]$ is precisely the set of all real-valued functions on X. In particular $\dim(\mathbb{R}[X]) = |X|$.

The **Lagrange interpolation functions** $\{p_z\}_{z \in X}$, defined by

$$p_z(x) := \begin{cases} 1, & \text{if } x = z \text{ and} \\ 0, & \text{if } x \neq z \end{cases}$$

are a basis for the vector space $\mathbb{R}[X]$.

1. Polynomial functions on X

Which functions $s : X \to \mathbb{R}$ are the restriction of a polynomial (i.e. live in $\mathbb{R}[X]$)?

Theorem.

If X is finite then $\mathbb{R}[X]$ is precisely the set of all real-valued functions on X. In particular $\dim(\mathbb{R}[X]) = |X|$.

The **Lagrange interpolation functions** $\{p_z\}_{z \in X}$, defined by

$$p_z(x) := \begin{cases} 1, & \text{if } x = z \text{ and} \\ 0, & \text{if } x \neq z \end{cases}$$

are a basis for the vector space $\mathbb{R}[X]$.
In the **Exercises** you will show that for finite X the Lagrange interpolation functions are the restrictions of polynomials.

1. Polynomial functions on X

Example:

Let $X := \{-1, 1\}^3$. *The coordinate ring*

$$\mathbb{R}[X] = \mathbb{R}[x, y, z]/(x^2 - 1, y^2 - 1, z^2 - 1)$$

and the square-free monomials $1, x, y, z, xy, xz, yz, xyz$ *are a basis for* $\mathbb{R}[X]$.

1. Polynomial functions on X

The ring $\mathbb{R}[X]$ is filtered by degree...

Definition.

Let r be a positive integer. A nonzero function $f \in \mathbb{R}[X]$ has degree $\leq r$ if it is the restriction to X of a polynomial h in $\mathbb{R}[x_1, \ldots, x_n]$ involving only monomials of degree $\leq r$.

In this case we write $f \in \mathbb{R}[X]_{\leq r}$.

- The sets $\mathbb{R}[X]_{\leq r}$ are vector spaces.
- We have $\mathbb{R}[X]_{\leq r} \cdot \mathbb{R}[X]_{\leq s} \subseteq \mathbb{R}[X]_{\leq s+t}$.

2. Nonnegative polynomials P

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

2. Nonnegative polynomials P

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

If $E \subseteq P$ is any collection of functions with a fast **membership algorithm** then we can use E to construct a lower bound on $f_{\min}$.

2. Nonnegative polynomials P

$$f_{\text{min}} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

If $E \subseteq P$ is any collection of functions with a fast **membership algorithm** then we can use E to construct a lower bound on f_{min}.

Lemma.

The inequality $f_{\text{min}} \geq \beta$ holds if

$$\beta := \max\{\lambda \in \mathbb{R} : f - \lambda \in E\}$$

2. Nonnegative polynomials P

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

If $E \subseteq P$ is any collection of functions with a fast **membership algorithm** then we can use E to construct a lower bound on $f_{\min}$.

Lemma.

The inequality $f_{\min} \geq \beta$ holds if

$$\beta := \max\{\lambda \in \mathbb{R} : f - \lambda \in E\}$$

Proof.

Since $E \subseteq P$ the feasible set of the new program is smaller. □

2. Nonnegative polynomials P

$$f_{\min} = \max\{\lambda \in \mathbb{R} : f - \lambda \in P\}$$

If $E \subseteq P$ is any collection of functions with a fast **membership algorithm** then we can use E to construct a lower bound on $f_{\min}$.

Lemma.

The inequality $f_{\min} \geq \beta$ holds if

$$\beta := \max\{\lambda \in \mathbb{R} : f - \lambda \in E\}$$

Proof.

Since $E \subseteq P$ the feasible set of the new program is smaller. □

Remark.

There are many natural sets $E \subseteq P$, leading to different approaches to optimization on X.

The SOS hierarchy [Parrilo, Lasserre, early 2000s]

Definition.

For each integer $r \geq 0$ define the cone of **sums of squares of functions of degree at most** r, *denoted $\Sigma_{\leq r}$ as*

$$\Sigma_{\leq r} := \left\{ g \in \mathbb{R}[X] : \exists m \in \mathbb{N}, s_1, \ldots, s_m \in \mathbb{R}[X]_{\leq r} \text{ with } g = \sum_{i=1}^{m} s_i^2 \right\}$$

Definition.

The **level** r **sum-of-squares bound** $f_{(r)}$ *for $f_{\min}$ is given by*

$$f_{(r)} := \max \{ \lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r} \}$$

The SOS hierarchy [Parrilo, Lasserre, early 2000s]

Definition.

For each integer $r \geq 0$ define the cone of **sums of squares of functions of degree at most** r, *denoted* $\Sigma_{\leq r}$ *as*

$$\Sigma_{\leq r} := \left\{ g \in \mathbb{R}[X] : \exists m \in \mathbb{N}, s_1, \ldots, s_m \in \mathbb{R}[X]_{\leq r} \text{ with } g = \sum_{i=1}^{m} s_i^2 \right\}$$

Definition.

The **level** r **sum-of-squares bound** $f_{(r)}$ *for* $f_{\min}$ *is given by*

$$f_{(r)} := \max \left\{ \lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r} \right\}$$

Since $\Sigma_{\leq r} \subseteq P$ *we have* $f_{(r)} \leq f_{\min}$.

Properties of $\Sigma_{\leq r}$ optimization

Whether $g \in \Sigma_{\leq r}$ can be reduced to a semidefinite programming feasibility problem. Computing $f_{(r)}$ is solving a semidefinite program.

Properties of $\Sigma_{\leq r}$ optimization

Whether $g \in \Sigma_{\leq r}$ can be reduced to a semidefinite programming feasibility problem. Computing $f_{(r)}$ is solving a semidefinite program.

Lemma.

If $\vec{m}$ is a vector whose components are a basis for $\mathbb{R}[X]_{\leq r}$ then

$$g \in \Sigma_{\leq r} \iff \exists A \left(A \succeq 0 \text{ and } g = \vec{m}^t A \vec{m} \text{ in } \mathbb{R}[X] \right)$$

Properties of $\Sigma_{\leq r}$ optimization

Whether $g \in \Sigma_{\leq r}$ can be reduced to a semidefinite programming feasibility problem. Computing $f_{(r)}$ is solving a semidefinite program.

Lemma.

If $\vec{m}$ is a vector whose components are a basis for $\mathbb{R}[X]_{\leq r}$ then

$$g \in \Sigma_{\leq r} \iff \exists A \left(A \succeq 0 \text{ and } g = \vec{m}^t A \vec{m} \text{ in } \mathbb{R}[X] \right)$$

Proof.

A matrix $A \succeq 0$ iff it admits a Choleski factorization $C^t C = A$.

Properties of $\Sigma_{\leq r}$ optimization

Whether $g \in \Sigma_{\leq r}$ can be reduced to a semidefinite programming feasibility problem. Computing $f_{(r)}$ is solving a semidefinite program.

Lemma.

If $\vec{m}$ is a vector whose components are a basis for $\mathbb{R}[X]_{\leq r}$ then

$$g \in \Sigma_{\leq r} \iff \exists A \left(A \succeq 0 \text{ and } g = \vec{m}^t A \vec{m} \text{ in } \mathbb{R}[X]\right)$$

Proof.

A matrix $A \succeq 0$ iff it admits a Choleski factorization $C^t C = A$.

$$g \in \Sigma_{\leq r} \iff g = \|C\vec{m}\|^2 = m^t C^t C m$$

Properties of $\Sigma_{\leq r}$ optimization

Whether $g \in \Sigma_{\leq r}$ can be reduced to a semidefinite programming feasibility problem. Computing $f_{(r)}$ is solving a semidefinite program.

Lemma.

If $\vec{m}$ is a vector whose components are a basis for $\mathbb{R}[X]_{\leq r}$ then

$$g \in \Sigma_{\leq r} \iff \exists A \left(A \succeq 0 \text{ and } g = \vec{m}^t A \vec{m} \text{ in } \mathbb{R}[X]\right)$$

Proof.

A matrix $A \succeq 0$ iff it admits a Choleski factorization $C^t C = A$.

$$g \in \Sigma_{\leq r} \iff g = \|C\vec{m}\|^2 = m^t C^t C m \iff g = \vec{m}^t A \vec{m} \text{ for } A \succeq 0$$

□

PSD matrices and SOS

If $\vec{m} = \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$ is a matrix with real entries then:

PSD matrices and SOS

If $\vec{m} = \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$ is a matrix with real entries then:

$$C\vec{m} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix} \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix} = \begin{pmatrix} x^2 + x + 1 \\ 2x^2 + 3x + 4 \\ 5x^2 + 6x + 7 \end{pmatrix}$$

PSD matrices and SOS

If $\vec{m} = \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix}$ and $C = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix}$ is a matrix with real entries then:

$$C\vec{m} = \begin{pmatrix} 1 & 1 & 1 \\ 2 & 3 & 4 \\ 5 & 6 & 7 \end{pmatrix} \begin{pmatrix} x^2 \\ x \\ 1 \end{pmatrix} = \begin{pmatrix} x^2 + x + 1 \\ 2x^2 + 3x + 4 \\ 5x^2 + 6x + 7 \end{pmatrix}$$

$$(x^2 + x + 1)^2 + (2x^2 + 3x + 4)^2 + (5x^2 + 6x + 7)^2 =$$
$$= \|C\vec{m}\|^2 = \vec{m}^t C^t C\vec{m} = \vec{m}^t A\vec{m}$$

Example: $\{-1, 1\}^3$

Let $f = 6 - 2(xy + xz + yz)$. Compute $f_{(1)}$.

Example: $\{-1,1\}^3$

Let $f = 6 - 2(xy + xz + yz)$. Compute $f_{(1)}$.

1. Compute a basis for $\mathbb{R}[X]_{\leq 1}$

Example: $\{-1,1\}^3$

Let $f = 6 - 2(xy + xz + yz)$. Compute $f_{(1)}$.

1. Compute a basis for $\mathbb{R}[X]_{\leq 1}$

$$\vec{m}^t = (1, x, y, z)$$

2. Maximize λ such that

$$6 - 2(xy + xz + yz) - \lambda = \vec{m}^t \begin{pmatrix} A & B & C & D \\ \cdot & E & F & G \\ \cdot & \cdot & H & I \\ \cdot & \cdot & \cdot & J \end{pmatrix} \vec{m} \text{ in R[X]}$$

and the matrix is PSD.

Example: $\{-1,1\}^3$

Let $f = 6 - 2(xy + xz + yz)$. Compute $f_{(1)}$.

1. Compute a basis for $\mathbb{R}[X]_{\leq 1}$

$$\vec{m}^t = (1, x, y, z)$$

2. Maximize λ such that

$$6 - 2(xy + xz + yz) - \lambda = \vec{m}^t \begin{pmatrix} A & B & C & D \\ \cdot & E & F & G \\ \cdot & \cdot & H & I \\ \cdot & \cdot & \cdot & J \end{pmatrix} \vec{m} \text{ in R[X]}$$

   and the matrix is PSD.

Maximize λ such that

$$(6-\lambda)-2(xy+xz+yz) = (A+E+H+J)+2(Bx+Cy+2D+Fxy+Gxz+Iyz)$$

and the matrix is PSD.

Example: $\{-1,1\}^3$

We conclude that if $f = 6 - 2(xy + xz + yz)$ then computing $f_{(1)}$ is equivalent to the following SDP:

$f_{(1)} := \max\,[6 - (A + E + H + J)]$ *subject to*

$$\begin{pmatrix} A & 0 & 0 & 0 \\ 0 & E & -1 & -1 \\ 0 & -1 & H & -1 \\ 0 & -1 & -1 & J \end{pmatrix} \succeq 0$$

Example: $\{-1, 1\}^3$

We conclude that if $f = 6 - 2(xy + xz + yz)$ then computing $f_{(1)}$ is equivalent to the following SDP:

$f_{(1)} := \max\left[6 - (A + E + H + J)\right]$ *subject to*

$$\begin{pmatrix} A & 0 & 0 & 0 \\ 0 & E & -1 & -1 \\ 0 & -1 & H & -1 \\ 0 & -1 & -1 & J \end{pmatrix} \succeq 0$$

Exercise

Prove that $f_{(1)} = 0$.

Natural questions for SOS hierarchy on finite sets.

Fix a finite set $X \subseteq \mathbb{R}^n$ and a function $f \in \mathbb{R}[X]_{\leq d}$.
For each integer r define

$$f_{(r)} = \max\left\{\lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r}\right\}.$$

We know that $f_{\min} \geq f_{(r)}$ for every r.

Natural questions for SOS hierarchy on finite sets.

Fix a finite set $X \subseteq \mathbb{R}^n$ and a function $f \in \mathbb{R}[X]_{\leq d}$.
For each integer r define

$$f_{(r)} = \max \{\lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r}\}.$$

We know that $f_{\min} \geq f_{(r)}$ for every r.

- *(SDP Size) What can we say about the size of the semidefinite programs needed to compute $f_{(r)}$?*

Natural questions for SOS hierarchy on finite sets.

Fix a finite set $X \subseteq \mathbb{R}^n$ and a function $f \in \mathbb{R}[X]_{\leq d}$.
For each integer r define

$$f_{(r)} = \max \{\lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r}\}.$$

We know that $f_{\min} \geq f_{(r)}$ for every r.

- *(SDP Size) What can we say about the size of the semidefinite programs needed to compute $f_{(r)}$?*
- *(Exactness degree) Is there a degree r such that the equality $f_{\min} = f_{(r)}$ holds for* ***every*** *f? If so what is the* ***minimal value*** *of r for which this happens?*

Natural questions for SOS hierarchy on finite sets.

Fix a finite set $X \subseteq \mathbb{R}^n$ and a function $f \in \mathbb{R}[X]_{\leq d}$.
For each integer r define

$$f_{(r)} = \max \{\lambda \in \mathbb{R} : f - \lambda \in \Sigma_{\leq r}\}.$$

We know that $f_{\min} \geq f_{(r)}$ for every r.

- *(SDP Size) What can we say about the size of the semidefinite programs needed to compute $f_{(r)}$?*
- *(Exactness degree) Is there a degree r such that the equality $f_{\min} = f_{(r)}$ holds for* **every** *f? If so what is the* **minimal value** *of r for which this happens?*
- *(Rate of convergence) Can we make* **quantitative estimates** *of how the worst-case gap $f_{\min} - f_{(r)}$ depends on r?*

Some good answers...

If $X = \{-1, 1\}^n \subseteq \mathbb{R}^n$ then we have complete answers (stated here for quadratic functions on $f \in \mathbb{R}[X]_{\leq 2}$ and n even)

Theorem.

The following statements hold:

1. *(Fawzi, Saunderson, Parrilo, 2016) If $r \geq n/2 + 1$ then $f_{\min} = f_{(r)}$.*
2. *(Blekherman, Gouveia, Pfeiffer, 2016) The above inequality is sharp.*
3. *(Laurent, Slot, 2021) If $\|f\|_\infty \leq 1$ then*

$$f_{\min} - f_{(r)} \leq C\zeta/n$$

where ζ is the largest root of a Krawtchouk polynomial.

Outline of the course

In this course we will try to understand the key ideas behind the proofs of the previous Theorems. We will focus on the case of the hypercube but the ideas can be applied to many other situations.

1. Lecture 1: The sum-of-squares hierarchy on finite sets.
2. Lecture 2: Upper bounds on exactness degree (**Sparsity**).
3. Lecture 3: Lower bounds on exactness degree (**Symmetry**).
4. Lecture 4: Quantitative bounds for $f_{\min} - f_{(r)}$ (**The polynomial Kernel method**).

Some commutative algebra:

The restriction map $\mathbb{R}[x_1, \ldots, x_n] \to \mathbb{R}[X]$ sends

$$\mathbb{R}[x_1, \ldots, x_n]_{\leq 0} \subset \mathbb{R}[x_1, \ldots, x_n]_{\leq 1} \subset \cdots \subset \mathbb{R}[x_1, \ldots, x_n]_{\leq r} \subset \ldots$$

$$\mathbb{R}[X]_{\leq 0} \subseteq \mathbb{R}[X]_{\leq 1} \subseteq \cdots \subseteq \mathbb{R}[X]_{\leq r} \subseteq \ldots$$

Definition.

The **Hilbert-Samuel function** *of X is*

$$h_X(r) := \dim\left(\mathbb{R}[X]_{\leq r}\right)$$

and the **Castelnuovo-Mumford regularity** *of X is*

$$\operatorname{reg}(X) := 1 + \min\{n \in \mathbb{N} : h_X(n) = |X|\}$$

Some commutative algebra:

The restriction map $\mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}[X]$ sends

$$\mathbb{R}[x_1, \dots, x_n]_{\leq 0} \subset \mathbb{R}[x_1, \dots, x_n]_{\leq 1} \subset \dots \subset \mathbb{R}[x_1, \dots, x_n]_{\leq r} \subset \dots$$

$$\mathbb{R}[X]_{\leq 0} \subseteq \mathbb{R}[X]_{\leq 1} \subseteq \dots \subseteq \mathbb{R}[X]_{\leq r} \subseteq \dots$$

Definition.

The **Hilbert-Samuel function** *of X is*

$$h_X(r) := \dim \left(\mathbb{R}[X]_{\leq r}\right)$$

and the **Castelnuovo-Mumford regularity** *of X is*

$$\mathrm{reg}(X) := 1 + \min\{n \in \mathbb{N} : h_X(n) = |X|\}$$

Exercise

Let $X = \{-1, 1\}^n \subseteq \mathbb{R}^n$. Compute $h_X(r)$ and verify that $\mathrm{reg}(X) = 1 + n$.

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \mathrm{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \operatorname{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

Proof.

(1) The size of the vector $\vec{m}$ equals $h_X(r)$ because its components are a basis of $\mathbb{R}[X]_{\leq r}$.

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \mathrm{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

Proof.

(1) The size of the vector $\vec{m}$ equals $h_X(r)$ because its components are a basis of $\mathbb{R}[X]_{\leq r}$.
(2) Let $(p_z)_{z \in X}$ be the Lagrange interpolation functions.

1. They satisfy $p_z^2 = p_z$,

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \mathrm{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

Proof.

(1) The size of the vector $\vec{m}$ equals $h_X(r)$ because its components are a basis of $\mathbb{R}[X]_{\leq r}$.
(2) Let $(p_z)_{z \in X}$ be the Lagrange interpolation functions.

1. They satisfy $p_z^2 = p_z$,
2. They have degree at most $\mathrm{reg}(X) - 1$.

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \mathrm{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

Proof.

(1) The size of the vector $\vec{m}$ equals $h_X(r)$ because its components are a basis of $\mathbb{R}[X]_{\leq r}$.
(2) Let $(p_z)_{z \in X}$ be the Lagrange interpolation functions.

1. They satisfy $p_z^2 = p_z$,
2. They have degree at most $\mathrm{reg}(X) - 1$.
3. If f is a nonnegative function on X then

$$f(x) = \sum_{y \in X} \left(\sqrt{f(y)} p_y(x) \right)^2$$

SDP size and exactness for finite X

Theorem.

1. *If $f \in \mathbb{R}[X]$ then $f_{(r)}$ can be computed from a semidefinite program on symmetric matrices of size $h_X(r) \times h_X(r)$.*
2. *If $r \geq \mathrm{reg}(X) - 1$ then $f_{\min} = f_{(r)}$ for every $f \in \mathbb{R}[X]$.*

Proof.

(1) The size of the vector $\vec{m}$ equals $h_X(r)$ because its components are a basis of $\mathbb{R}[X]_{\leq r}$.
(2) Let $(p_z)_{z \in X}$ be the Lagrange interpolation functions.

1. They satisfy $p_z^2 = p_z$,
2. They have degree at most $\mathrm{reg}(X) - 1$.
3. If f is a nonnegative function on X then

$$f(x) = \sum_{y \in X} \left(\sqrt{f(y)} p_y(x) \right)^2$$

Outline of the course

In this course we will try to understand the key ideas behind the key results of SOS optimization on hypercubes. The basic ideas underlying them can be applied to many other situations.

1. Lecture 1: The sum-of-squares hierarchy on finite sets.
2. Lecture 2: Upper bounds on exactness degree (**Sparsity**).
3. Lecture 3: Lower bounds on exactness degree (**Symmetry**).
4. Lecture 4: Quantitative bounds for $f_{\min} - f_{(r)}$ (**The polynomial Kernel method**).