# Semidefinite programming hierarchies for polynomial programs

CMU 10-725 Optimization
Fall 2012

Shiva Kaul <skkaul@cs.cmu.edu>

(yes, this is going to be handwritten.)

There's life beyond convex optimization.

(Now you tell us...)

# Example 1: Sparse estimation

## Example 1: Sparse estimation

Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n$, $\; y_t = \langle w^*, x_t \rangle + \varepsilon_t$

Goal: recover $w^*$

## Example 1: Sparse estimation

Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n$, $\; y_t = \langle w^*, x_t \rangle + \varepsilon_t$

Goal: recover $w^*$

Lasso: $$\min_{w} \; \sum_{t=1}^{T} \left( y_t - \langle w, x_t \rangle \right)^2 + \lambda \sum_{j=1}^{n} |w_j|$$

## Example 1: Sparse estimation

Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n$, $y_t = \langle w^*, x_t \rangle + \varepsilon_t$

Goal: recover $w^*$

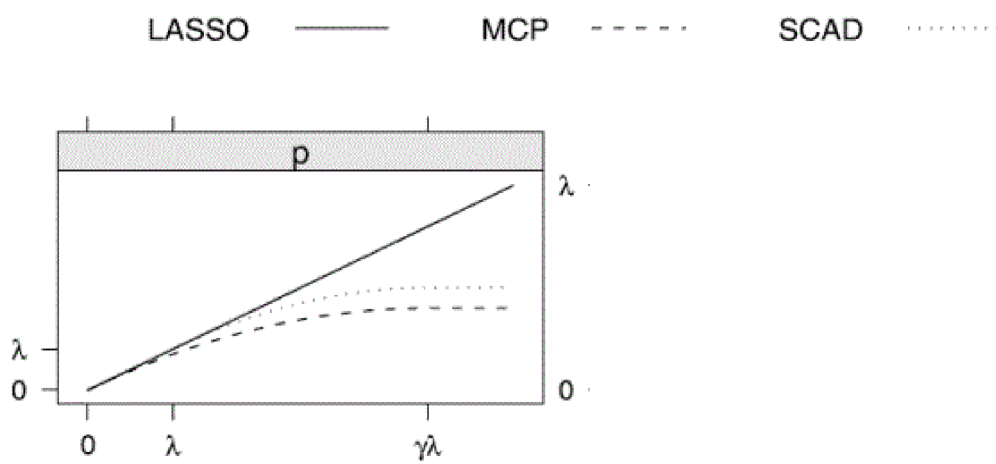
Lasso: $\displaystyle \min_{w} \sum_{t=1}^{T} \left(y_t - \langle w, x_t \rangle\right)^2 + \lambda \sum_{j=1}^{n} |w_j|$

Statistically, not an ideal estimator. [Zha12]

- if $\min_j |w_j^*|$ is large, $\|w - w^*\|$ is suboptimal by a factor of $\sqrt{\ln(n)}$ ("Lasso bias")
- under "$\ell_2$ regularity conditions", $w$ does not have the correct support (even asymptotically)

MCP penalty: $\min_{w} \sum_{t=1}^{T} (y_t - \langle w, x_t \rangle)^2 + \lambda \sum_{j=1}^{n} \rho(|w_j|)$

$$\rho(|w_j|) = \begin{cases} \lambda |w_j| - w_j^2/2\gamma & |w_j| \le \lambda\gamma \\ \frac{1}{2}\gamma\lambda^2 & \text{otherwise} \end{cases}$$



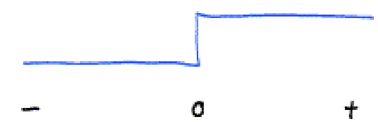

[Bre11]

# Example 2: Linear Classification

# Example 2: Linear Classification

Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n, \quad y_t \in \{-1, 1\}$ from distribution $D$

Goal: $\min\limits_{w} \mathop{\mathbb{E}}\limits_{(x,y)\sim D} \left[ \mathbb{1}(\langle w, x_t \rangle \neq y_t) \right]$

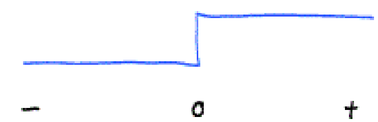



0-1 loss

# Example 2: Linear Classification

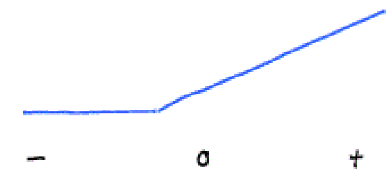
Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n, \quad y_t \in \{-1, 1\}$ from distribution $D$

Goal: $\min\limits_{w} \mathop{\mathbb{E}}\limits_{(x,y)\sim D} \left[ \mathbb{1}(\langle w, x_t \rangle \neq y_t) \right]$




0-1 loss

SVM: $\min\limits_{w} \max(0, 1 - y_t \langle w, x_t \rangle) + \lambda \|w\|_2^2$




hinge loss

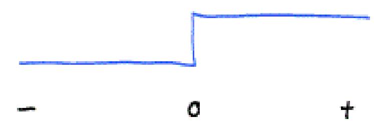
Relation to original problem is hazy

- need margin or exponential amount of data [Sha10]

# Example 2: Linear Classification

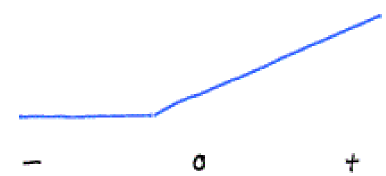
Given: $\{(x_t, y_t)\}_{t=1}^{T}$ $\quad x_t \in \mathbb{R}^n, \quad y_t \in \{-1, 1\}$ from distribution $\mathcal{D}$

Goal: $\min_{w} \; \mathbb{E}_{(x,y)\sim \mathcal{D}} \left[ \mathbb{1}(\langle w, x_t \rangle \neq y_t) \right]$



0-1 loss

SVM: $\min_{w} \; \max(0, 1 - y_t \langle w, x_t \rangle) + \lambda \|w\|_2^2$



hinge loss

Relation to original problem is hazy

- need margin or exponential amount of data [Sha10]

Classical learning theory supports regularized ERM:

$$\min_{w} \; \sum_{t=1}^{T} \mathbb{1}\left[ \operatorname{sgn}(\langle w, x_t \rangle) \neq y_t \right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

# Polynomial optimization

# Polynomial optimization

$$\min.\ p(x) \quad \text{s.t.} \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

where $p$ and $g_i$ are polynomials

- in $n$ variables $x_1, \ldots, x_n$
- of degree $d$
- with rational coefficients

# Polynomial optimization

$$\min. \quad p(x) \qquad \text{s.t.} \qquad g_1(x) \geq 0, \; \ldots, \; g_m(x) \geq 0$$

where $p$ and $g_i$ are polynomials

- in $n$ variables $x_1, \ldots, x_n$
- of degree $d$
- with rational coefficients

- Very general.

# Polynomial optimization

$$\min.\ p(x) \quad \text{s.t.} \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

where $p$ and $g_i$ are polynomials

- in $n$ variables $x_1, \ldots, x_n$
- of degree $d$
- with rational coefficients

- Very general.
- Not always well-defined.

# Classification as a polynomial program:

## Classification as a polynomial program:

$$\min_{w} \ \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}\left(\langle w, x_t \rangle\right) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} \ell_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\ell_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} \ell_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\left[\ \ell_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}\right.$$

$$\ell_t \in \{0, 1\}$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} \ell_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\left[\ \ell_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}\right.$$

$$\ell_t \in \{0, 1\} \longleftrightarrow \ell_t = \ell_t^2$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} l_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\left[\, l_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}\right.$$

$$l_t \in \{0,1\} \longleftrightarrow l_t = l_t^2 \longleftrightarrow \begin{array}{l} l_t^2 - l_t \geq 0 \\ l_t - l_t^2 \geq 0 \end{array}$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t\rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} \ell_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\left[\ \ell_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}\right.$$

$$\ell_t \in \{0,1\} \longleftrightarrow \ell_t = \ell_t^2 \longleftrightarrow \begin{array}{l} \ell_t^2 - \ell_t \geq 0 \\ \ell_t - \ell_t^2 \geq 0 \end{array}$$

$$-y_t \langle w, x_t \rangle \leq \ell_t \cdot B$$

## Classification as a polynomial program:

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad \|w\|_2^2 \leq B$$

$$\min_{w} \sum_{t=1}^{T} \mathbb{1}\left[\operatorname{sgn}(\langle w, x_t \rangle) \neq y_t\right] \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\min_{w} \sum_{t=1}^{T} \ell_t \quad \text{s.t.} \quad B^2 - \sum_i w_i^2 \geq 0$$

$$\left[\ \ell_t = \begin{cases} 1 & w \text{ incorrect on } x_t \\ 0 & \text{otherwise} \end{cases}\right.$$

$$\ell_t \in \{0, 1\} \longleftrightarrow \ell_t = \ell_t^2 \longleftrightarrow \begin{matrix} \ell_t^2 - \ell_t \geq 0 \\ \ell_t - \ell_t^2 \geq 0 \end{matrix}$$

$$-y_t \langle w, x_t \rangle \leq \ell_t \cdot B$$

$$\|w\|_2 \geq \varepsilon$$

## polynomials as vectors (in basis of monomials)

## polynomials as vectors (in basis of monomials)

$$p(x) = 5 + 2x_1 + 3x_2 - x_1x_2 + 4.7x_2^2$$

## polynomials as vectors (in basis of monomials)

$$p(x) = 5 + 2x_1 + 3x_2 - x_1x_2 + 4.7x_2^2$$

$$= \langle (5, 2, 3, -1, 0, 4.7), (x_1^0x_2^0, x_1^1x_2^0, x_1^0x_2^1, x_1^1x_2^1, x_1^2x_2^0, x_1^0x_2^2) \rangle$$

$(5, 2, 3, -1, 0, 4.7) \longrightarrow \vec{p}$ (will drop $\rightarrow$)

$\phi_D(x)$ is the vector of monomials of $x$ up to degree $D$.

of dimension $\binom{n+D}{D}$ (here, $D = 2$)

## polynomials as vectors (in basis of monomials)

$$p(x) = 5 + 2x_1 + 3x_2 - x_1x_2 + 4.7x_2^2$$

$$= \langle \underbrace{(5,\ 2,\ 3,\ -1,\ 0,\ 4.7)}_{\to\ \vec{p}},\ (x_1^0x_2^0,\ x_1^1x_2^0,\ x_1^0x_2^1,\ x_1^1x_2^1,\ x_1^2x_2^0,\ x_1^0x_2^2)\rangle$$

(will drop $\rightarrow$)

$\phi_D(x)$ is the vector of monomials of $x$ up to degree $D$.

of dimension $\binom{n+D}{D}$ (here, $D = 2$)

$$= \sum_{\substack{\alpha = (\alpha_1, \dots, \alpha_D) \\ \text{s.t. } |\alpha| \le D}} \vec{p}_\alpha \, x^\alpha \quad \text{where} \quad x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$$

(here, $n = 2$)

# semidefinite programming

# semidefinite programming

$$\min_{y} \ \langle c, y \rangle \quad \text{s.t.} \quad \sum_{\gamma} y_\gamma A^i_\gamma \succeq 0 \quad \forall i, \quad Cy = b$$

linear objective

PSD cone constraints

affine constraints

# semidefinite programming

$$\min_{y} \; \langle c, y \rangle \quad \text{s.t.} \quad \sum_{\gamma} y_\gamma A^i_\gamma \succeq 0 \;\; \forall i, \quad Cy = b$$

linear objective

PSD cone constraints

affine constraints

one PSD constraint

infinitely many linear constraints

$$P \succeq 0 \quad \Longleftrightarrow \quad v^T P v \geq 0 \;\; \forall v$$

## semidefinite programming

$$\min_{y} \ \langle c, y\rangle \quad \text{s.t.} \quad \sum_{\gamma} y_\gamma A^i_\gamma \succeq 0 \quad \forall i, \quad Cy = b$$

linear objective

PSD cone constraints

affine constraints

one PSD constraint

infinitely many linear constraints

$$P \succeq 0 \quad \Longleftrightarrow \quad v^T P v \geq 0 \quad \forall v$$

## Encoding the constraints

PSD cone is pointed:

$$P \succeq 0 \ \wedge \ -P \succeq 0 \ \Rightarrow \ P = 0$$

## semidefinite programming

$$\min_{y} \langle c, y \rangle \quad \text{s.t.} \quad \sum_{\gamma} y_\gamma A^i_\gamma \succeq 0 \quad \forall i, \quad Cy = b$$

linear objective

PSD cone constraints

affine constraints

one PSD constraint

infinitely many linear constraints

$$P \succeq 0 \iff v^T P v \geq 0 \quad \forall v$$

## Encoding the constraints

PSD cone is pointed:

$$P \succeq 0 \wedge -P \succeq 0 \Rightarrow P = 0$$

Given $P \succeq 0$, $P \neq 0$

$$g_i(x) \geq 0 \iff g_i(x) P \succeq 0$$

$0 \preceq$

$$0 \preceq \phi_D(x)\, \phi_D(x)^T$$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta}
\end{aligned}
$$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta} \\
&= \left[x^\alpha x^\beta\right]_{\alpha,\beta}
\end{aligned}
$$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta} \\
&= \left[x^\alpha\,x^\beta\right]_{\alpha,\beta} \qquad x^\alpha x^\beta = x^\gamma \quad \text{for} \quad \alpha+\beta=\gamma
\end{aligned}
$$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta} \\
&= \left[x^\alpha x^\beta\right]_{\alpha,\beta} \\
&= \sum_\gamma x^\gamma I_\gamma
\end{aligned}
$$

$$
\begin{array}{ccc}
 & & \beta \\
 & & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\alpha & \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{bmatrix} & & & 1 & & \\ & & 1 & & & \\ & 1 & & & & \\ 1 & & & & & \\ & & & & & \end{bmatrix}
\end{array}
\qquad
\begin{array}{l} n = 1 \\ \gamma = 3 \end{array}
$$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$0 \preceq \phi_D(x)\,\phi_D(x)^T$$

$$= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta}$$

$$= \left[x^\alpha x^\beta\right]_{\alpha,\beta}$$

$$= \sum_\gamma x^\gamma I_\gamma$$

$$\begin{array}{cc} & \beta \\ \alpha & \begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & & & & 1 & & \\ 1 & & & 1 & & & \\ 2 & & 1 & & & & \\ 3 & 1 & & & & & \\ 4 & & & & & & \end{array} \end{array}$$

$n = 1$
$\gamma = 3$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$0 \preceq g_i(x)\,\phi_D(x)\,\phi_D(x)^T \iff g_i(x) \geq 0$$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha\, \phi_D(x)_\beta\right]_{\alpha,\beta} \\
&= \left[x^\alpha\, x^\beta\right]_{\alpha,\beta} \\
&= \sum_\gamma x^\gamma I_\gamma
\end{aligned}
$$

$$
\beta
$$

$$
\begin{array}{cc}
 & \begin{array}{cccccc} 0 & 1 & 2 & 3 & 4 & 5 \end{array} \\
\alpha \;\begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \\ 4 \end{array} & \left[\begin{array}{cccccc} & & & 1 & & \\ & & 1 & & & \\ & 1 & & & & \\ 1 & & & & & \\ & & & & & \end{array}\right]
\end{array}
$$

$n = 1$

$\gamma = 3$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$
0 \preceq g_i(x)\,\phi_D(x)\,\phi_D(x)^T \iff g_i(x) \geq 0
$$

$$
= \left(\sum_\alpha g^i_\alpha x^\alpha\right)\left(\sum_\beta I_\beta\, x^\beta\right)
$$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T$$

$$= \left[\phi_D(x)_\alpha\, \phi_D(x)_\beta\right]_{\alpha,\beta}$$

$$= \left[x^\alpha x^\beta\right]_{\alpha,\beta}$$

$$= \sum_\gamma x^\gamma I_\gamma$$

$$\begin{array}{c|c} & \beta \\ \alpha & \begin{array}{c|cccccc} & 0 & 1 & 2 & 3 & 4 & 5 \\ \hline 0 & & & & 1 & & \\ 1 & & & 1 & & & \\ 2 & & 1 & & & & \\ 3 & 1 & & & & & \\ 4 & & & & & & \end{array} \end{array}$$

$n = 1$

$\gamma = 3$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$0 \preceq g_i(x)\,\phi_D(x)\,\phi_D(x)^T \iff g_i(x) \geq 0$$

$$= \left(\sum_\alpha g^i_\alpha x^\alpha\right)\left(\sum_\beta I_\beta\, x^\beta\right)$$

$$= \sum_\alpha \sum_\beta g^i_\alpha I_\beta\, x^{\alpha+\beta}$$

$$
\begin{array}{c|cccccc}
 & \beta & & & & & \\
\alpha & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
0 & & & & 1 & & \\
1 & & & 1 & & & \\
2 & & 1 & & & & \\
3 & 1 & & & & & \\
4 & & & & & & 
\end{array}
$$

$n = 1$
$\gamma = 3$

$$
\begin{aligned}
0 &\preceq \phi_D(x)\,\phi_D(x)^T \\
&= \left[\phi_D(x)_\alpha \, \phi_D(x)_\beta\right]_{\alpha,\beta} \\
&= \left[x^\alpha x^\beta\right]_{\alpha,\beta} \\
&= \sum_\gamma x^\gamma I_\gamma
\end{aligned}
$$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$0 \preceq g_i(x)\,\phi_D(x)\,\phi_D(x)^T \iff g_i(x) \geq 0$$

$$
\begin{aligned}
&= \left(\sum_\alpha g^i_\alpha x^\alpha\right)\left(\sum_\beta I_\beta x^\beta\right) \\
&= \sum_\alpha \sum_\beta g^i_\alpha I_\beta \, x^{\alpha+\beta} \\
&= \sum_\gamma x^\gamma \sum_{\alpha+\beta=\gamma} g^i_\alpha I_\beta
\end{aligned}
$$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T$$

$$= \left[\phi_D(x)_\alpha\,\phi_D(x)_\beta\right]_{\alpha,\beta}$$

$$= \left[x^\alpha x^\beta\right]_{\alpha,\beta}$$

$$= \sum_\gamma x^\gamma I_\gamma$$

$$\begin{array}{cc|cccccc|}
 & & \multicolumn{6}{c}{\beta} \\
 & & 0 & 1 & 2 & 3 & 4 & 5 \\
\hline
 & 0 & & & & 1 & & \\
 & 1 & & & 1 & & & \\
\alpha & 2 & & 1 & & & & \\
 & 3 & 1 & & & & & \\
 & 4 & & & & & & \\
\hline
\end{array}$$

$n = 1$
$\gamma = 3$

$x^\alpha x^\beta = x^\gamma$ for $\alpha + \beta = \gamma$

$I_\gamma$ is matrix with 1s in these cells.

$$0 \preceq g_i(x)\,\phi_D(x)\,\phi_D(x)^T \qquad \Leftrightarrow \qquad g_i(x) \geq 0$$

$$= \left(\sum_\alpha g^i_\alpha x^\alpha\right)\left(\sum_\beta I_\beta x^\beta\right)$$

$$= \sum_\alpha \sum_\beta g^i_\alpha I_\beta\, x^{\alpha+\beta}$$

$$= \sum_\gamma x^\gamma \underbrace{\sum_{\alpha+\beta=\gamma} g^i_\alpha I_\beta}_{A^i_\gamma}$$

## The relaxation

$$\min_{x} \sum_{\gamma} p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_{\gamma} x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

## The relaxation

$$\min_{x} \sum_{\gamma} p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_{\gamma} x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$\vdots$

linearize $x^\alpha$ to $y_\alpha$

$\vdots$

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

The relaxation

$$\min_{x} \sum_{\gamma} p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_{\gamma} x^\gamma A^i_\gamma \geq 0 \quad \forall 1 \leq i \leq m$$

⋮

linearize $x^\alpha$ to $y_\alpha$

⋮

$$\min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \leq D} y_\gamma A^i_\gamma \geq 0 \quad \forall 1 \leq i \leq m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad y_0 = 1$

## The relaxation

$$\min_{x} \sum_{\gamma} p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_{\gamma} x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$\vdots$

linearize $x^\alpha$ to $y_\alpha$

$\vdots$

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad y_0 = 1$

$0 \preceq \phi_D(x)\, \phi_D(x)^T$

## The relaxation

$$\min_x \sum_\gamma p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_\gamma x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

⋮

linearize $x^\alpha$ to $y_\alpha$

⋮

$$\min_y \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad y_0 = 1$

$$\begin{aligned} 0 &\preceq \phi_D(x)\,\phi_D(x)^T \\ &= \left[x^\alpha x^\beta\right]_{|\alpha|,|\beta| \le D} \\ &= \left[x^{\alpha+\beta}\right]_{|\alpha|,|\beta| \le D} \end{aligned}$$

## The relaxation

$$\min_x \sum_\gamma p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_\gamma x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

⋮

linearize $x^\alpha$ to $y_\alpha$

⋮

$$\min_y \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad y_0 = 1$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T = \left[x^\alpha x^\beta\right]_{|\alpha|,|\beta| \le D} = \left[x^{\alpha+\beta}\right]_{|\alpha|,|\beta| \le D}$$

$$\left[y_{\alpha+\beta}\right]_{|\alpha|,|\beta| \le D} \succeq 0$$

## The relaxation

$$\min_x \sum_\gamma p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_\gamma x^\gamma A^i_\gamma \succeq 0 \quad \forall 1 \le i \le m$$

⋮

linearize $x^\alpha$ to $y_\alpha$

⋮

$$\min_y \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad y_0 = 1$

$$0 \preceq \phi_D(x)\phi_D(x)^T = [x^\alpha x^\beta]_{|\alpha|,|\beta| \le D} = [x^{\alpha+\beta}]_{|\alpha|,|\beta| \le D}$$

free

$$[y_{\alpha+\beta}]_{|\alpha|,|\beta| \le D} \succeq 0$$

## The relaxation

$$\min_x \sum_\gamma p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_\gamma x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

linearize $x^\alpha$ to $y_\alpha$

$$\min_y \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$

$y_0 = 1$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T = \left[x^\alpha x^\beta\right]_{|\alpha|,|\beta| \le D} = \left[x^{\alpha+\beta}\right]_{|\alpha|,|\beta| \le D}$$

free

$$\left[y_{\alpha+\beta}\right]_{|\alpha|,|\beta| \le D} \succeq 0$$

- not free, need constraint

## The relaxation

$$\min_{x} \sum_{\gamma} p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_{\gamma} x^\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

linearize $x^\alpha$ to $y_\alpha$

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall\, 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$ $\qquad$ $y_0 = 1$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T = [x^\alpha x^\beta]_{|\alpha|,|\beta| \le D} = [x^{\alpha+\beta}]_{|\alpha|,|\beta| \le D}$$

free

$$[y_{\alpha+\beta}]_{|\alpha|,|\beta| \le D} \succeq 0$$

- not free, need constraint
- what $D$? mantra: the more constraints, the tighter the relaxation

## The relaxation

$$\min_x \sum_\gamma p_\gamma x^\gamma \quad \text{s.t.} \quad \sum_\gamma x^\gamma A^i_\gamma \succeq 0 \quad \forall 1 \le i \le m$$

$\vdots$

linearize $x^\alpha$ to $y_\alpha$

$\vdots$

$$\min_y \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0 \quad \forall 1 \le i \le m$$

$x^0 = 1$. Could "cheat" by taking $y_0 \to \pm\infty$

$$y_0 = 1$$

$$0 \preceq \phi_D(x)\,\phi_D(x)^T$$
$$= [x^\alpha x^\beta]_{|\alpha|,|\beta| \le D}$$
$$= [x^{\alpha+\beta}]_{|\alpha|,|\beta| \le D}$$

free

$$\Updownarrow$$

$$g_0(x) = 1, \; 0 \preceq g_0(x)\,\phi_D(x)\,\phi_D(x)^T$$

$$[y_{\alpha+\beta}]_{|\alpha|,|\beta| \le D} \succeq 0$$

- not free, need constraint
- what $D$? mantra: the more constraints, the tighter the relaxation

# Moment relaxation of order $D$

$$v_{pop} = \min_{x} \quad p(x) \quad \text{s.t.} \quad 1 = g_0(x) \geq 0, \; g_1(x) \geq 0, \; \dots, \; g_m(x) \geq 0$$

$$v_{mom}^{D} = \min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall\, 0 \leq i \leq m$$

$$y_0 = 1$$

where $A_\gamma^i = \sum_{\alpha+\beta=\gamma} g_\alpha^i I_\beta$

$I_\beta$ has 1 at $(a,b)$ iff $a+b=\beta$

# Moment relaxation of order $D$

$$v_{pop} = \min_{x} \; p(x) \quad \text{s.t.} \quad 1 = g_0(x) \geq 0, \; g_1(x) \geq 0, \; \dots, \; g_m(x) \geq 0$$

$$v_{mom}^{D} = \min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall\, 0 \leq i \leq m$$

$$y_0 = 1$$

where $$A_\gamma^i = \sum_{\alpha+\beta=\gamma} g_\alpha^i I_\beta$$

$I_\beta$ has 1 at $(a,b)$ iff $a+b=\beta$

- $\binom{n+D}{D}$ variables

# Moment relaxation of order $D$

$$v_{pop} = \min_{x} \; p(x) \quad \text{s.t.} \quad 1 = g_0(x) \geq 0, \; g_1(x) \geq 0, \; \ldots, \; g_m(x) \geq 0$$

$$v_{mom}^{D} = \min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall\, 0 \leq i \leq m$$

$$y_0 = 1$$

where $A_\gamma^i = \sum_{\alpha + \beta = \gamma} g_\alpha^i I_\beta$

$I_\beta$ has 1 at $(a,b)$ iff $a + b = \beta$

- $\binom{n+D}{D}$ variables
- Lasserre hierarchy: $v_{mom}^{1} \leq v_{mom}^{2} \leq \cdots \leq v_{pop}$

# Moment relaxation of order D

$$v_{pop} = \min_{x} \; p(x) \quad \text{s.t.} \quad 1 = g_0(x) \geq 0, \; g_1(x) \geq 0, \; \dots, \; g_m(x) \geq 0$$

$$v_{mom}^{D} = \min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad \sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall\, 0 \leq i \leq m$$

$$y_0 = 1$$

where $A_\gamma^i = \sum_{\alpha + \beta = \gamma} g_\alpha^i I_\beta$

$I_\beta$ has 1 at $(a,b)$ iff $a+b = \beta$

- $\binom{n+D}{D}$ variables
- Lasserre hierarchy: $v_{mom}^1 \leq v_{mom}^2 \leq \dots \leq v_{pop}$
- Under some conditions, for big enough D, it is <u>tight</u>!

$v_{mom}^{D^*} = v_{pop}$

## In the dual: sums of squares

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2$$ where $s_j$ is a polynomial

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

# In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

$$s_j(x)^2 = \langle s, \phi(x)\rangle \langle s, \phi(x)\rangle$$

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

$$\begin{aligned} s_j(x)^2 &= \langle s, \phi(x)\rangle \langle s, \phi(x)\rangle \\ &= (s_j^T \phi(x))^T (s_j^T \phi(x)) \end{aligned}$$

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

$$\begin{aligned} s_j(x)^2 &= \langle s, \phi(x)\rangle \langle s, \phi(x)\rangle \\ &= (s_j^T \phi(x))^T (s_j^T \phi(x)) \\ &= \phi(x)^T s_j s_j^T \phi(x) \end{aligned}$$

## In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

$$s_j(x)^2 = \langle s, \phi(x)\rangle \langle s, \phi(x)\rangle$$

$$= (s_j^T \phi(x))^T (s_j^T \phi(x))$$

$$= \phi(x)^T s_j s_j^T \phi(x)$$

$$\sum_j s_j(x)^2 = \phi(x)^T Q \phi(x) \qquad Q = \sum_j s_j s_j^T$$

# In the dual: sums of squares

$$q(x) = \sum_{j=1}^{k} s_j(x)^2 \quad \text{where } s_j \text{ is a polynomial}$$

clearly nonnegative

convex!

Every SOS is a quadratic form in the monomials of $x$:

$$\begin{aligned} s_j(x)^2 &= \langle s, \phi(x)\rangle \langle s, \phi(x)\rangle \\ &= (s_j^T \phi(x))^T (s_j^T \phi(x)) \\ &= \phi(x)^T s_j s_j^T \phi(x) \end{aligned}$$

$$\sum_j s_j(x)^2 = \phi(x)^T Q \, \phi(x) \qquad Q = \sum_j s_j s_j^T$$

of appropriate size

Conversely, every $Q \succeq 0$ defines a sum of squares:

$$Q = S^T S = \begin{bmatrix} | & & | \\ s_1 & \cdots & s_k \\ | & & | \end{bmatrix} \begin{bmatrix} — & s_1 & — \\ & \vdots & \\ — & s_k & — \end{bmatrix}$$

# Abstraction

Abstraction

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \qquad \text{s.t.} \quad y_0 = 1$$

$$\sum_{|\gamma| \le D} y_\gamma A^i_\gamma \ge 0 \qquad \forall\, 0 \le i \le m$$

Abstraction

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \qquad \text{s.t.} \quad y_0 = 1$$

$$\underbrace{\sum_{|\gamma| \le D} y_\gamma A^i_\gamma}_{L_i y} \ge 0 \qquad \forall\, 0 \le i \le m$$

$$L_i y \ge 0$$

# Abstraction

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad y_0 = 1$$

$$\underbrace{\sum_{|\gamma| \le D} y_\gamma A^i_\gamma \succeq 0}_{g_i^T y \, y y^T \;=\; L_i y \;\succeq 0} \quad \forall\, 0 \le i \le m$$

$$\boxed{\begin{aligned} & g_i(x)\, \phi_D(x)\, \phi_D(x)^T \\ &= \Big(\sum_\alpha g^i_\alpha x^\alpha\Big)\Big(\sum_\beta I_\beta x^\beta\Big) \\ &= \sum_\alpha \sum_\beta g^i_\alpha I_\beta \cdot x^{\alpha+\beta} \\ &= \sum_\gamma x^\gamma \underbrace{\sum_{\alpha+\beta=\gamma} g^i_\alpha I_\beta}_{A^i_\gamma} \end{aligned}}$$

# Taking the dual

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

Which is the dual of:

$$\max_{\ell} \langle b, \ell \rangle \quad \text{s.t.} \quad p - C^T \ell \in K$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

Which is the dual of:

$$\max_{l} \langle b, l \rangle \quad \text{s.t.} \quad p - C^T l \in K$$

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdot \\ 0 & 0 & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} = C^T, \quad \text{so} \quad p - C^T l = p - \langle (1, 0, \ldots), l \rangle$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

Which is the dual of:

$$\max_{\ell} \langle b, \ell \rangle \quad \text{s.t.} \quad p - C^T \ell \in K$$

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdot \\ 0 & 0 & 0 & \vdots \\ 0 & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} = C^T, \quad \text{so} \quad p - C^T \ell = p - \langle (1, 0, \ldots), \ell \rangle$$

$$= (p_0 - \ell_0, \ldots\ldots)$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \le d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \ge 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \qquad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

Which is the dual of:

$$\max_{l} \langle b, l \rangle \qquad \text{s.t.} \quad p - C^T l \in K$$

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdot \\ 0 & 0 & 0 & \cdot \\ 0 & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} = C^T, \quad \text{so} \quad p - C^T l = p - \langle (1, 0, \ldots), l \rangle$$

$$= (p_0 - l_0, \ldots\ldots)$$

$$\equiv p(\cdot) - l_0$$

## Taking the dual

$$\min_{y} \sum_{|\gamma| \leq d} p_\gamma y_\gamma \quad \text{s.t.} \quad L_i y \geq 0 \quad \forall 0 \leq i \leq m, \quad y_0 = 1$$

Let's write it in the following form:

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^*, \quad Cy = b$$

Which is the dual of:

$$\max_{l} \langle b, l \rangle \quad \text{s.t.} \quad p - C^T l \in K$$

$$C = \begin{bmatrix} 1 & 0 & 0 & \cdot \\ 0 & 0 & 0 & \vdots \\ 0 & \vdots & \vdots & \vdots \\ \cdot & \cdot & \cdot & \cdot \end{bmatrix} = C^T, \quad \text{so} \quad p - C^T l = p - \langle (1, 0, \ldots), l \rangle$$

$$= (p_0 - l_0, \ldots\ldots)$$

$$b = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix}, \quad \text{so} \quad \langle b, l \rangle = l_0 \qquad \equiv p(\cdot) - l_0$$

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad L_i y \geq 0 \quad \forall 0 \leq i \leq m, \quad y_0 = 1$$

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \quad , \quad y_0 = 1$$

$$\max_{l_0} l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \in K$$

<u>The dual cone</u>

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad L_i y \geq 0 \quad \forall 0 \leq i \leq m \,, \quad y_0 = 1$$

$$\min_{y} \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \qquad , \quad y_0 = 1$$

$$\max_{l_0} \; l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \in K$$

## The dual cone

$$L_i y \geq 0 \quad \forall 0 \leq i \leq m$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad L_i y \succeq 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \quad , \quad y_0 = 1$$

$$\max_{l_0} \ l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \ \in K$$

<u>The dual cone</u>

$$L_i y \succeq 0 \quad \forall\, 0 \le i \le m$$

$$\equiv \ y \in L_i^{-1} Z \quad \forall\, 0 \le i \le m \qquad Z = \text{PSD cone}$$

$$\min_y \ \langle p, y \rangle \quad \text{s.t.} \quad L_i y \succeq 0 \quad \forall\, 0 \le i \le m \,, \quad y_0 = 1$$

$$\min_y \ \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \,, \quad y_0 = 1$$

$$\max_{l_0} \ l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \in K$$

<u>The dual cone</u>

$$L_i y \succeq 0 \quad \forall\, 0 \le i \le m$$

$$\equiv y \in L_i^{-1} Z \quad \forall\, 0 \le i \le m \qquad Z = \text{PSD cone}$$

$$\equiv y \in \bigcap_{i=0}^{m} L_i^{-1} Z$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad L_i y \succeq 0 \quad \forall\, 0 \le i \le m, \quad y_0 = 1$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \quad , \quad y_0 = 1$$

$$\max_{l_0} \ l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \in K$$

<u>The dual cone</u>

$$L_i y \succeq 0 \quad \forall\, 0 \le i \le m$$

$$\equiv y \in L_i^{-1} Z \quad \forall\, 0 \le i \le m \qquad Z = \text{PSD cone}$$

$$\equiv y \in \underbrace{\bigcap_{i=0}^{m} \underbrace{L_i^{-1} Z}_{K_i^*}}_{K^*}$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad L_i\, y \geq 0 \quad \forall\, 0 \leq i \leq m \ , \quad y_0 = 1$$

$$\min_{y} \ \langle p, y \rangle \quad \text{s.t.} \quad y \in K^* \quad , \quad y_0 = 1$$

$$\max_{l_0} \ l_0 \quad \text{s.t.} \quad p(\cdot) - l_0 \ \in K$$

<u>The dual cone</u>

$$L_i\, y \geq 0 \quad \forall\, 0 \leq i \leq m$$

$$\equiv y \in L_i^{-1} Z \quad \forall\, 0 \leq i \leq m \qquad Z = \text{PSD cone}$$

$$\equiv y \in \underbrace{\bigcap_{i=0}^{m} \underbrace{L_i^{-1} Z}_{K_i^*}}_{K^*}$$

$$K = K_0 + \dots + K_m = \left\{ \sum_{i=0}^{m} R_i \ : \ R_i \in K_i \right\}$$

$\mathbb{R}^n$ $\qquad$ $\mathbb{R}^{\binom{n+D}{D}}$ $\qquad$ $\left(\mathbb{R}^{\binom{n+D}{D}}\right)^2$

$x$ $\qquad$ $\phi(x)$ $\qquad$ $L_i(\phi(x))$

$L_i(y)$


$L_i(K_i^*) \subseteq Z = Z^*$

$q_i : \mathbb{R}^n \to \mathbb{R}$

$q_i(x) = \langle Q_i , L_i(\phi(x)) \rangle$

For $Q_i \in Z$

$\langle Q_i, P \rangle \geq 0 \quad \forall \; P \in Z$

$\langle Q_i, P \rangle \geq 0 \quad \forall \; P \in L_i K_i^*$

$\langle Q_i, L_i y \rangle \geq 0 \quad \forall \; y \in K_i^*$

$\langle L_i^T Q , y \rangle \geq 0 \quad \forall \, y \in K_i^*$

$\to$ subset of $K_i$

# What is $K_i$ ?

# What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

# What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \qquad \forall\, Q \succeq 0 \}$$

## What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

## What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \in Z \}$$

# What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \in Z \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, L_i^T Q \in L_i^T Z \}$$

# What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \in Z \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, L_i^T Q \in L_i^T Z \}$$

$$= \{ \langle y, R \rangle \geq 0 \quad \forall\, R \in L_i^T Z \}$$

## What is $K_i$ ?

$$K_i^* = \{ y : L_i y \succeq 0 \}$$

$$= \{ \langle L_i y, Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \succeq 0 \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, Q \in Z \}$$

$$= \{ \langle y, L_i^T Q \rangle \geq 0 \quad \forall\, L_i^T Q \in L_i^T Z \}$$

$$= \{ \langle y, R \rangle \geq 0 \quad \forall\, R \in \underbrace{L_i^T Z}_{K_i} \}$$

## A functional view of $R \in K_i = L_i^T Z$

$R(y) =$

# A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \geq 0$$

# A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \geq 0$$

$$= \langle L_i y, Q \rangle$$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \geq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \; y y^T, Q \rangle$$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \succeq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \, \langle y y^T, Q \rangle$$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \succeq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \langle y y^T, Q \rangle$$

$$= g_i^T y \, \mathrm{tr}(y y^T Q)$$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \succeq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \langle y y^T, Q \rangle$$

$$= g_i^T y \, tr(y y^T Q)$$

$$= g_i^T y \, tr(y^T Q y)$$

# A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \geq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \, \langle y y^T, Q \rangle$$

$$= g_i^T y \, tr(y y^T Q)$$

$$= g_i^T y \, tr(y^T Q y)$$

$$= g_i^T y \; y^T Q y \quad \text{for some } Q \geq 0$$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \succeq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \, \langle y y^T, Q \rangle$$

$$= g_i^T y \, \mathrm{tr}(y y^T Q)$$

$$= g_i^T y \, \mathrm{tr}(y^T Q y)$$

$$= g_i^T y \; y^T Q y \qquad \text{for some } Q \succeq 0$$

So $K = K_0 + \cdots + K_m = \left\{ \sum_{i=0}^{m} R_i \ : \ R_i \in K_i \right\}$

## A functional view of $R \in K_i = L_i^T Z$

$$R(y) = \langle y, R \rangle = \langle y, L_i^T Q \rangle \quad \text{for some } Q \succeq 0$$

$$= \langle L_i y, Q \rangle$$

$$= \langle g_i^T y \, y y^T, Q \rangle$$

$$= g_i^T y \, \langle y y^T, Q \rangle$$

$$= g_i^T y \, \mathrm{tr}(y y^T Q)$$

$$= g_i^T y \, \mathrm{tr}(y^T Q y)$$

$$= g_i^T y \, y^T Q y \qquad \text{for some } Q \succeq 0$$

So $K = K_0 + \cdots + K_m = \left\{ \sum_{i=0}^{m} R_i \; : \; R_i \in K_i \right\}$

$$\equiv \left\{ y \mapsto \sum_{i=0}^{m} g_i^T y \, y^T Q_i y \; : \; Q_i \succeq 0 \right\}$$

## But what is $K$, really?

## But what is K, really?

$$K = \left\{ y \mapsto \sum_{i=0}^{m} g_i^T y \; y^T Q_i \, y \qquad : \qquad Q_i \succeq 0 \right\}$$

## But what is K, really?

$$K = \left\{ y \mapsto \sum_{i=0}^{m} g_i^T y \; y^T Q_i \, y \quad : \quad Q_i \succeq 0 \right\}$$

$$= \left\{ x \mapsto \sum_{i=0}^{m} g_i(x)\, q_i(x) \quad : \quad q_i \text{ is a sum of squares of degree } D \right\}$$

# But what is K, really?

$$K = \left\{ y \mapsto \sum_{i=0}^{m} g_i^T y \; y^T Q_i y \quad : \quad Q_i \succeq 0 \right\}$$

$$= \left\{ x \mapsto \sum_{i=0}^{m} g_i(x)\, q_i(x) \quad : \quad q_i \text{ is a sum of squares of degree } D \right\}$$

$$= \left\{ x \mapsto q_0(x) + \sum_{i=1}^{m} g_i(x)\, q_i(x) \quad : \quad q_i \text{ is a sum of squares of degree } D \right\}$$

## But what is K, really?

$$K = \left\{ y \mapsto \sum_{i=0}^{m} g_i^T y \, y^T Q_i y \quad : \quad Q_i \succeq 0 \right\}$$

$$= \left\{ x \mapsto \sum_{i=0}^{m} g_i(x)\, q_i(x) \quad : \quad q_i \text{ is a sum of squares of degree } D \right\}$$

$$= \left\{ x \mapsto q_0(x) + \sum_{i=1}^{m} g_i(x)\, q_i(x) \quad : \quad q_i \text{ is a sum of squares of degree } D \right\}$$

AKA the truncated quadratic module

to degree D

## The SDP hierarchies

$$v_{pop} = \min.\ p(x) \quad \text{s.t.} \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

$$v_{mom}^D = \min_{y}.\ \langle p, y \rangle \quad \text{s.t.} \quad [y_\alpha y_\beta]_{|\alpha|, |\beta| \leq D} \succeq 0$$

$$\sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall 1 \leq i \leq m$$

$$y_0 = 1$$

$$v_{sos}^D = \max_{l_0}\ l_0 \quad \text{s.t.} \quad p - l_0 = q_0 + \sum_{i=1}^{m} g_i q_i$$

where $q_i$ is a sum of squares of degree $D$

# The SDP hierarchies

$$V_{pop} = \min. \quad p(x) \quad \text{s.t.} \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

$$V_{mom}^{D} = \min_{y} \langle p, y \rangle \quad \text{s.t.} \quad [y_\alpha y_\beta]_{|\alpha|, |\beta| \leq D} \succeq 0$$

$$\sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall 1 \leq i \leq m$$

$$y_0 = 1$$

$$V_{sos}^{D} = \max_{l_0} \quad l_0 \quad \text{s.t.} \quad p - l_0 = q_0 + \sum_{i=1}^{m} g_i q_i$$

where $q_i$ is a sum of squares of degree $D$

$$\left.\begin{array}{l} V_{mom}^{1} \leq V_{mom}^{2} \leq \cdots \\ V_{sos}^{1} \leq V_{sos}^{2} \leq \cdots \end{array}\right\} \leq V_{pop}$$

## The SDP hierarchies

$$V_{pop} = \min.\ p(x) \quad \text{s.t.} \quad g_1(x) \geq 0, \ldots, g_m(x) \geq 0$$

$$V_{mom}^D = \min_y.\ \langle p, y \rangle \quad \text{s.t.} \quad [y_\alpha y_\beta]_{|\alpha|, |\beta| \leq D} \succeq 0$$

$$\sum_{|\gamma| \leq D} y_\gamma A_\gamma^i \succeq 0 \quad \forall 1 \leq i \leq m$$

$$y_0 = 1$$

$$V_{sos}^D = \max_{l_0}\ l_0 \quad \text{s.t.} \quad p - l_0 = q_0 + \sum_{i=1}^m g_i q_i$$

where $q_i$ is a sum of squares of degree $D$

$$\begin{array}{ccccc} V_{mom}^1 & \leq & V_{mom}^2 & \leq & \cdots \\ \text{VI} & & \text{VI} & & \\ V_{sos}^1 & \leq & V_{sos}^2 & \leq & \cdots \end{array} \leq V_{pop}$$

by weak duality.
(Strong duality under usual conditions)

# Intuition behind the SOS relaxation

## Intuition behind the SOS relaxation

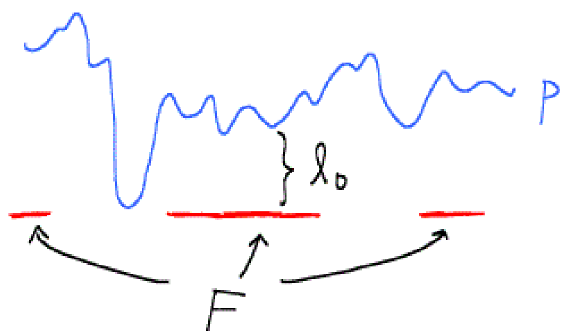
$$\min_{x \in F} \; p(x) \qquad F = \{ x : g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \}$$

# Intuition behind the SOS relaxation

$$\min_{x \in F} \; p(x) \qquad F = \{ x : g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \}$$

$\updownarrow$ equivalent

$$\max_{\ell_0} \; \ell_0 \quad \text{s.t.} \quad p(x) - \ell_0 \geq 0 \quad \forall x \in F$$



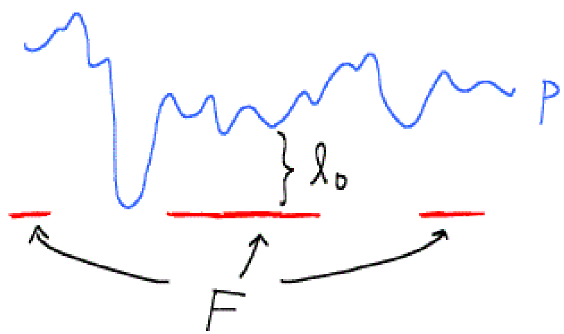

how much can we lower $p$ while maintaining nonnegativity on $F$?

# Intuition behind the SOS relaxation

$$\min_{x \in F} \; p(x) \qquad F = \{ x : g_1(x) \geq 0, \ldots, g_m(x) \geq 0 \}$$

$\updownarrow$ equivalent

$$\max_{l_0} \; l_0 \quad \text{s.t.} \quad p(x) - l_0 \geq 0 \quad \forall x \in F$$




how much can we lower $p$ while maintaining nonnegativity on $F$?

$$\max_{l_0} \; l_0 \quad \text{s.t.}$$

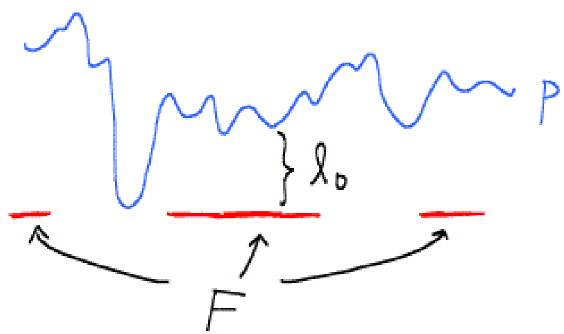
$$p - l_0 = q_0 + \sum_{i=1}^{m} g_i q_i \quad \text{where } q_i \text{ is SOS of degree } D$$

# Intuition behind the SOS relaxation

$$\min_{x \in F} \; p(x) \qquad F = \{ x : g_1(x) \geq 0, \dots, g_m(x) \geq 0 \}$$

$\Updownarrow$ equivalent

$$\max_{l_0} \; l_0 \quad \text{s.t.} \quad p(x) - l_0 \geq 0 \quad \forall x \in F$$




how much can we lower $p$ while maintaining nonnegativity on $F$?

$$\max_{l_0} \; l_0 \quad \text{s.t.}$$

$$p - l_0 = q_0 + \sum_{i=1}^{m} g_i q_i \quad \text{where } q_i \text{ is SOS of degree } D$$

$\Rightarrow$ $p - l_0$ has a certificate of nonnegativity (at round $D$)

# When do SOS certificates exist?

## When do SOS certificates exist?

If $F \subseteq \{x : \|x\|_2 \leq B\}$ (i.e. some $g_i(x) = \sum_{i=1}^{n} x_i^2 - B^2$)

## When do SOS certificates exist?

If $F \subseteq \{x : \|x\|_2 \le B\}$ (i.e. some $g_i(x) = \sum_{i=1}^{n} x_i^2 - B^2$)

then as $D \to \infty$,

if $p > 0$ on $F$, then $p \in K$

## When do SOS certificates exist?

If $F \subseteq \{x : \|x\|_2 \leq B\}$ (i.e. some $g_i(x) = \sum_{i=1}^{n} x_i^2 - B^2$)

then as $D \to \infty$,

if $p > 0$ on $F$, then $p \in K$

(that is, $V_{sos}^{D} \to V_{pop}$)

## When do SOS certificates exist?

If $F \subseteq \{x : \|x\|_2 \leq B\}$ (i.e. some $g_i(x) = \sum_{i=1}^{n} x_i^2 - B^2$)

then as $D \to \infty$,

if $p > 0$ on $F$, then $p \in K$

(that is, $V_{sos}^{D} \to V_{pop}$)

This is Putinar's Positivstellensatz, a specialization of a central result in algebraic geometry.

## In practice: exploiting problem structure

# In practice: exploiting problem structure

- Extracting minimizing $x$ from MOM solution $y$

  Putinar's Psatz
  $p$ has a unique minimum $\Rightarrow$ $x_i = x^{e_i} = y_{e_i}$
  $D$ large enough

## In practice: exploiting problem structure

- Extracting minimizing $x$ from MOM solution $y$

  Putinar's Psatz
  $p$ has a unique minimum $\Rightarrow$ $x_i = x^{e_i} = y_{e_i}$
  $D$ large enough

- Hierarchy convergence for finite $D$
  - in practice, low $D$ is fine

# In practice: exploiting problem structure

- Extracting minimizing $x$ from MOM solution $y$

  Putinar's Psatz
  $p$ has a unique minimum $\implies$ $x_i = x^{e_i} = y_{e_i}$
  $D$ large enough

- Hierarchy convergence for finite $D$
  - in practice, low $D$ is fine

- Approximation quality as $D$ grows

## In practice: exploiting problem structure

- Extracting minimizing $x$ from MOM solution $y$

  Putinar's Psatz
  $p$ has a unique minimum $\Rightarrow \quad x_i = x^{e_i} = y_{e_i}$
  $D$ large enough

- Hierarchy convergence for finite $D$
  - in practice, low $D$ is fine
- Approximation quality as $D$ grows
- Sparsity

Go forth and optimize!