# Inverse polynomial optimization

Jean B. Lasserre

LAAS-CNRS and Institute of Mathematics, Toulouse, France

Workshop: Conic optimization: IMS Singapore, November 2012

- semidefinite programming
- Inverse polynomial optimization
- A hierarchy of semidefinite programs:
- The canonical “sparse" form of an optimal solution
- a by-product

- semidefinite programming
- Inverse polynomial optimization
- A hierarchy of semidefinite programs:
- The canonical “sparse" form of an optimal solution
- a by-product

- semidefinite programming
- Inverse polynomial optimization
- A hierarchy of semidefinite programs:
- The canonical “sparse" form of an optimal solution
- a by-product

- semidefinite programming
- Inverse polynomial optimization
- A hierarchy of semidefinite programs:
- The canonical “sparse" form of an optimal solution
- a by-product

- semidefinite programming
- Inverse polynomial optimization
- A hierarchy of semidefinite programs:
- The canonical “sparse" form of an optimal solution
- a by-product

$$\mathbf{P} \quad \rightarrow \quad \min_{x \in \mathbb{R}^n} \{ c' x \mid \sum_{i=1}^{n} A_i x_i \succeq b \},$$

$$\mathbf{P}^* \quad \rightarrow \quad \max_{Y \in \mathcal{S}_m} \{ \langle b, Y \rangle \mid \quad Y \succeq 0; \langle A_i, Y \rangle = c_i, \quad i = 1, \ldots, n \}$$

- $c \in \mathbb{R}^n$ and $b, A_i, Y \in \mathcal{S}_m$ ($m \times m$ symmetric matrices)
- $Y \succeq 0$ means $Y$ semidefinite positive; $\langle A, B \rangle = \mathrm{trace}\,(AB)$.

**P** and its dual $\mathbf{P}^*$ are **convex** problems that are solvable in polynomial time to arbitrary precision $\epsilon > 0$.

= generalization to the convex cone $\mathcal{S}_m^+$ ($X \succeq 0$) of Linear Programming on the convex polyhedral cone $\mathbb{R}_+^m$ ($x \geq 0$).

# Semidefinite Programming

$$\mathbf{P} \quad \to \quad \min_{x \in \mathbb{R}^n} \{ c' x \mid \sum_{i=1}^{n} A_i x_i \succeq b \},$$

$$\mathbf{P}^* \quad \to \quad \max_{Y \in \mathcal{S}_m} \{ \langle b, Y \rangle \mid \; Y \succeq 0; \; \langle A_i, Y \rangle = c_i, \quad i = 1, \ldots, n \}$$

- $c \in \mathbb{R}^n$ and $b, A_i, Y \in \mathcal{S}_m$ ($m \times m$ symmetric matrices)
- $Y \succeq 0$ means $Y$ semidefinite positive; $\langle A, B \rangle = \mathrm{trace}\,(AB)$.

**P** and its dual $\mathbf{P}^*$ are **convex** problems that are solvable in polynomial time to arbitrary precision $\epsilon > 0$.
= generalization to the convex cone $\mathcal{S}_m^+$ ($X \succeq 0$) of Linear Programming on the convex polyhedral cone $\mathbb{R}_+^m$ ($x \geq 0$).

- weak duality: $\langle b, Y \rangle \leq c' x$ for all feasible $x \in \mathbb{R}^n$, $Y \in \mathcal{S}_m$.
- strong duality: under "Slater interior point condition"

$$\exists x \in \mathbb{R}^n,\ Y \succ 0; \quad \sum_{i=1}^{n} A_i x_i \succ b; \quad \langle A_i, Y \rangle = c_i \quad i = 1, \ldots, n.$$

Then there is **no duality gap** and

$$\sup \mathbf{P}^* = \max \mathbf{P}^* = \min \mathbf{P} = \inf \mathbf{P}^*$$

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages. Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

- weak duality: $\langle b, Y\rangle \leq c' x$ for all feasible $x \in \mathbb{R}^n$, $Y \in \mathcal{S}_m$.
- strong duality: under "Slater interior point condition"

$$\exists x \in \mathbb{R}^n,\ Y \succ 0; \quad \sum_{i=1}^n A_i x_i \succ b; \quad \langle A_i, Y\rangle = c_i \quad i = 1, \ldots, n.$$

Then there is **no duality gap** and

$$\sup \mathbf{P}^* = \max \mathbf{P}^* = \min \mathbf{P} = \inf \mathbf{P}^*$$

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages. Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

- weak duality: $\langle b, Y\rangle \leq c' x$ for all feasible $x \in \mathbb{R}^n$, $Y \in \mathcal{S}_m$.
- strong duality: under "Slater interior point condition"

$$\exists x \in \mathbb{R}^n, \; Y \succ 0; \quad \sum_{i=1}^{n} A_i x_i \succ b; \quad \langle A_i, Y\rangle = c_i \quad i = 1, \ldots, n.$$

Then there is **no duality gap** and

$$\sup \mathbf{P}^* = \max \mathbf{P}^* = \min \mathbf{P} = \inf \mathbf{P}^*$$

Several academic SDP software packages exist, (e.g. MATLAB "LMI toolbox", SeduMi, SDPT3, ...). However, so far, size limitation is more severe than for LP software packages. Pioneer contributions by A. Nemirovsky, Y. Nesterov, N.Z. Shor, B.D. Yudin,...

Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial and

$$\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n \; : \; g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\},$$

for some polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$.

**... and consider the polynomial optimization problem:**

$$\mathbf{P}: \qquad f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \}$$

**What is the associated inverse optimization problem?**

Let $f \in \mathbb{R}[\mathbf{x}]$ be a polynomial and

$$\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n \;:\; g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\},$$

for some polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$.

**... and consider the polynomial optimization problem:**

$$\mathbf{P}: \qquad f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) \;:\; \mathbf{x} \in \mathbf{K} \}$$

**What is the associated inverse optimization problem?**

Given $\mathbf{y} \in \mathbf{K}$, one searches for a polynomial $g^* \in \mathbb{R}[\mathbf{x}]$,

AS CLOSE AS POSSIBLE to $f$,

and such that

**... $y$ is a global optimal solution of**

$$\min_{\mathbf{x}} \{ g^*(\mathbf{x}) \ : \ \mathbf{x} \in \mathbf{K} \}$$

i.e., $g^*(y) = \min_{\mathbf{x}} \{ g^*(\mathbf{x}) \ : \ \mathbf{x} \in \mathbf{K} \}$, AND SO ....

**the inverse optimization problem associated with P and $y$ reads:**

$$\mathbf{P}^{-1} : \quad \min_{g \in \mathbb{R}[\mathbf{x}]} \{ \| f - g \| \ : \ g(\mathbf{x}) - g(\mathbf{y}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{K} \}$$

for some appropriate norm $\| \cdot \|$ on $\mathbb{R}[\mathbf{x}]$.

Given $\mathbf{y} \in \mathbf{K}$, one searches for a polynomial $g^* \in \mathbb{R}[\mathbf{x}]$,

AS CLOSE AS POSSIBLE to $f$,

and such that

... $y$ is a global optimal solution of

$$\min_{\mathbf{x}} \{g^*(\mathbf{x}) \;:\; \mathbf{x} \in \mathbf{K}\}$$

i.e., $g^*(y) = \min_{\mathbf{x}} \{g^*(\mathbf{x}) \;:\; \mathbf{x} \in \mathbf{K}\}$, AND SO ....

the inverse optimization problem associated with **P** and $y$ reads:

$$\mathbf{P}^{-1}: \quad \min_{g \in \mathbb{R}[\mathbf{x}]} \{\|f - g\| \;:\; g(\mathbf{x}) - g(\mathbf{y}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{K}\}$$

for some appropriate norm $\|\cdot\|$ on $\mathbb{R}[\mathbf{x}]$.

Given $\mathbf{y} \in \mathbf{K}$, one searches for a polynomial $g^* \in \mathbb{R}[\mathbf{x}]$,

AS CLOSE AS POSSIBLE to $f$,

and such that

**... $y$ is a global optimal solution of**

$$\min_{\mathbf{x}} \{ g^*(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \}$$

i.e., $g^*(y) = \min_{\mathbf{x}} \{ g^*(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \}$, AND SO ....

**the inverse optimization problem associated with P and $y$ reads:**

$$\mathbf{P}^{-1} : \quad \min_{g \in \mathbb{R}[\mathbf{x}]} \{ \| f - g \| \; : \; g(\mathbf{x}) - g(\mathbf{y}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{K} \}$$

for some appropriate norm $\| \cdot \|$ on $\mathbb{R}[\mathbf{x}]$.

In general it makes sense to search for a polynomial $g$ of same degree as $f$, but not necessarily.

## Flexibility

- One may add structural constraints on $g$. For instance, writing $f$ in the canonical basis of monomials, $\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, one may impose the structural constraint $g_\alpha = 0$ whenever $f_\alpha = 0$, to obtain a polynomial with same "pattern".
- One may impose $g$ to be convex on $\mathbf{K}$ by imposing

$$y^T \nabla^2 g(\mathbf{x}) \, y \geq 0, \quad \forall \mathbf{x} \in \mathbf{K}, \, \forall y \in \{z \, : \, \|z\|^2 \leq 1\}.$$

In general it makes sense to search for a polynomial $g$ of same degree as $f$, but not necessarily.

## Flexibility

- One may add structural constraints on $g$. For instance, writing $f$ in the canonical basis of monomials, $\mathbf{x} \mapsto f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, one may impose the structural constraint $g_\alpha = 0$ whenever $f_\alpha = 0$, to obtain a polynomial with same "pattern".

- One may impose $g$ to be convex on $\mathbf{K}$ by imposing

$$y^T \nabla^2 g(\mathbf{x})\, y \geq 0, \quad \forall \mathbf{x} \in \mathbf{K},\, \forall\, y \in \{z \,:\, \|z\|^2 \leq 1\}.$$

# Motivation

## I. Practical ...

e.g., suppose that $\mathbf{y} \in \mathbf{K}$ is the $n$-th iterate of some local minimization algorithm. Then a practical issue is:

Why spend more energy (and computation) to find a (global?) minimum $\mathbf{x}^* \in \mathbf{K}$? whereas ...

- $f$ is perhaps not the "real" criterion .. just one among many other possibilities, and
- $\mathbf{y}$ could be an optimal solution of another criterion $g$ "close" to $f$!

# Motivation

## I. Practical ...

e.g., suppose that $\mathbf{y} \in \mathbf{K}$ is the $n$-th iterate of some local minimization algorithm. Then a practical issue is:

Why spend more energy (and computation) to find a (global?) minimum $\mathbf{x}^* \in \mathbf{K}$? whereas ...

- $f$ is perhaps not the "real" criterion .. just one among many other possibilities, and
- $\mathbf{y}$ could be an optimal solution of another criterion $g$ "close" to $f$!

# Motivation

## I. Practical ...

e.g., suppose that $\mathbf{y} \in \mathbf{K}$ is the $n$-th iterate of some local minimization algorithm. Then a practical issue is:

Why spend more energy (and computation) to find a (global?) minimum $\mathbf{x}^* \in \mathbf{K}$? whereas ...

- $f$ is perhaps not the "real" criterion .. just one among many other possibilities, and
- $\mathbf{y}$ could be an optimal solution of another criterion $g$ "close" to $f$!

# Motivation (continued)

## II. Mathematical ...

- If $\mathbf{y} \in \mathbf{K}$ is "close" to an optimal solution of $\mathbf{P}$, and $g^* \in \mathbb{R}[\mathbf{x}]$ solves the inverse optimization problem $\mathbf{P}^{-1}$, then

  $\|f - g^*\|$ is a measure of sensitivity or a kind of condition number on problem $\mathbf{P}$:
  The smaller $\|f - g^*\|$ is, the less sensitive to data is $\mathbf{P}$.

- If $\mathbf{y} \in \mathbf{K}$ is an optimal solution of $\mathbf{P}$ but not certified, then $\|f - g^*\|$ measures how hard it is to certify that $\mathbf{y}$ is optimal for $\mathbf{P}$.

# Motivation (continued)

## II. Mathematical ...

- If $\mathbf{y} \in \mathbf{K}$ is "close" to an optimal solution of **P**, and $g^* \in \mathbb{R}[\mathbf{x}]$ solves the inverse optimization problem $\mathbf{P}^{-1}$, then

  $\|f - g^*\|$ is a measure of sensitivity or a kind of condition number on problem **P**:
  The smaller $\|f - g^*\|$ is, the less sensitive to data is **P**.

- If $\mathbf{y} \in \mathbf{K}$ is an optimal solution of **P** but not certified, then $\|f - g^*\|$ measures how hard it is to certify that $\mathbf{y}$ is optimal for **P**.

# Solving the inverse optimization problem $\mathbf{P}^{-1}$

Let $d \geq \mathrm{deg} f$ and recall the inverse optimization problem:

$$\mathbf{P}^{-1}: \quad \min_{g \in \mathbb{R}[\mathbf{x}]_d} \{\|f - g\| \,:\, g(\mathbf{x}) - g(\mathbf{y}) \geq 0, \quad \forall \mathbf{x} \in \mathbf{K}\}$$

(and possibly additional structural constraints on $g$).

**Lemma**

*Let $\mathbf{K} \subset \mathbb{R}^n$ have a nonempty interior. The inverse problem $\mathbf{P}^{-1}$ has an optimal solution $g^* \in \mathbb{R}[\mathbf{x}]_d$.*

**To solve $\mathbf{P}^{-1}$ practically ... the difficulty is**

to express in a tractable manner that $\mathbf{y}$ is an optimal solution of

$$\min_{\mathbf{x}} \{ g^*(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \}$$

i.e., $g^*(\mathbf{x}) - g^*(\mathbf{y}) \geq 0$ for all $\mathbf{x} \in \mathbf{K}$.

**To solve $\mathbf{P}^{-1}$ practically ... the difficulty is**

to express in a tractable manner that $\mathbf{y}$ is an optimal solution of

$$\min_{\mathbf{x}} \{g^*(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K}\}$$

i.e., $g^*(\mathbf{x}) - g^*(\mathbf{y}) \geq 0$ for all $\mathbf{x} \in \mathbf{K}$.

This is why previous work has considered LPs, or some particular combinatorial problems. E.g., Burton and Toint (shortest path problems), Ahuja and Orlin (LPs), and Schaefer (Integer Programming).

For instance, in IP, the characterization by Schaefer is exponential in the input size of the problem and not practical.

# The inverse optimization problem $\mathbf{P}^{-1}$ (continued)

**However, for Polynomial Optimization ...**

and this is the main message to retain ...

- CERTIFICATES of global optimality EXIST!,
  e.g., Schmüdgen's and Putinar's Positivstellensätze.
- They can be translated into LMIs (or feasible solutions of semidefinite programs)!
- The SIZE of the certificate can be adjusted (to some extent), according to the computational workload limitation

# The inverse optimization problem $\mathbf{P}^{-1}$ (continued)

**However, for Polynomial Optimization ...**

and this is the main message to retain ...

- CERTIFICATES of global optimality EXIST!, e.g., Schmüdgen's and Putinar's Positivstellensätze.
- They can be translated into LMIs (or feasible solutions of semidefinite programs)!
- The SIZE of the certificate can be adjusted (to some extent), according to the computational workload limitation

# The inverse optimization problem $\mathbf{P}^{-1}$ (continued)

**However, for Polynomial Optimization ...**

and this is the main message to retain ...

- CERTIFICATES of global optimality EXIST!, e.g., Schmüdgen's and Putinar's Positivstellensätze.
- They can be translated into LMIs (or feasible solutions of semidefinite programs)!
- The SIZE of the certificate can be adjusted (to some extent), according to the computational workload limitation

# The inverse optimization problem $\mathbf{P}^{-1}$ (continued)

**However, for Polynomial Optimization ...**

and this is the main message to retain ...

- CERTIFICATES of global optimality EXIST!, e.g., Schmüdgen's and Putinar's Positivstellensätze.
- They can be translated into LMIs (or feasible solutions of semidefinite programs)!
- The SIZE of the certificate can be adjusted (to some extent), according to the computational workload limitation

# The inverse optimization problem $\mathbf{P}^{-1}$ (continued)

**Putinar's certificate for $\mathbf{P}^{-1}$**

Let $g \in \mathbb{R}[\mathbf{x}]_d$ for some $d \in \mathbb{N}$, and with $k \in \mathbb{N}$ fixed, replace

$$g(\mathbf{x}) - g(\mathbf{y}) \geq 0 \quad \forall \mathbf{x} \in \mathbf{K}, \quad \text{with}$$

$$g(\mathbf{x}) - g(\mathbf{y}) = \underbrace{\sigma_0(\mathbf{x})}_{\text{sos of deg } 2k} + \sum_{j=1}^{m} g_j(\mathbf{x}) \times \underbrace{\sigma_j(\mathbf{x})}_{\text{sos of deg } 2(k-v_j)}$$

for all $\mathbf{x} \in \mathbb{R}^n$.

The SOS polynomials $(\sigma_j)$ provide a Putinar's certificate that $y$ is a global minimizer of $g$ on $\mathbf{K}$!

Similarly ....if one searches for a polynomial $g$ convex on $\mathbf{K}$, it suffices to add the constraint:

$$y^T \nabla^2 g(\mathbf{x})\, y \;=\; \underbrace{\psi_0(\mathbf{x}, y)}_{SOS} + \sum_{j=1}^{m} \underbrace{\psi_j(\mathbf{x}, y)}_{SOS}\, g_j(\mathbf{x}) + \underbrace{\psi_{m+1}(\mathbf{x}, y)}_{S0S}\,(1 - \|y\|^2).$$

# A rationale for Putinar's certificate

## Why introduce this positivity certificate ?

Let $\mathbf{K} := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\}$ be compact and assume that the quadratic polynomial $\mathbf{x} \mapsto N - \|\mathbf{x}\|^2$ satisfies:

$$N - \|\mathbf{x}\|^2 = p_0 + \sum_{j=1}^{m} p_j \, g_j,$$

for some SOS polynomials $(p_j) \subset \mathbb{R}[\mathbf{x}]$.

**Theorem (Putinar's Positivstellensatz)**

*If $f \in \mathbb{R}[\mathbf{x}]$ is positive on $\mathbf{K}$ then:*

$$f = \sigma_0 + \sum_{j=1}^{m} \sigma_j \, g_j,$$

*for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.*

## Why introduce this positivity certificate ?

Let $\mathbf{K} := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\}$ be compact and assume that the quadratic polynomial $\mathbf{x} \mapsto N - \|\mathbf{x}\|^2$ satisfies:

$$N - \|\mathbf{x}\|^2 = p_0 + \sum_{j=1}^{m} p_j \, g_j,$$

for some SOS polynomials $(p_j) \subset \mathbb{R}[\mathbf{x}]$.

**Theorem (Putinar's Positivstellensatz)**

*If $f \in \mathbb{R}[\mathbf{x}]$ is positive on $\mathbf{K}$ then:*

$$f = \sigma_0 + \sum_{j=1}^{m} \sigma_j \, g_j,$$

*for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.*

## Why introduce this positivity certificate ?

Let $\mathbf{K} := \{\mathbf{x} : g_j(\mathbf{x}) \geq 0, \quad j = 1, \ldots, m\}$ be compact and assume that the quadratic polynomial $\mathbf{x} \mapsto N - \|\mathbf{x}\|^2$ satisfies:

$$N - \|\mathbf{x}\|^2 = p_0 + \sum_{j=1}^{m} p_j \, g_j,$$

for some SOS polynomials $(p_j) \subset \mathbb{R}[\mathbf{x}]$.

**Theorem (Putinar's Positivstellensatz)**

*If $f \in \mathbb{R}[\mathbf{x}]$ is positive on $\mathbf{K}$ then:*

$$f \;=\; \sigma_0 + \sum_{j=1}^{m} \sigma_j \, g_j,$$

*for some SOS polynomials $(\sigma_j) \subset \mathbb{R}[\mathbf{x}]$.*

**And in fact,**

from recent results by Marshall (2009) and Nie (2012) ...

**Putinar's Positivstellensatz**

also holds generically for polynomials of degree $d$ nonnegative on $\mathbf{K}$

## And in fact,

from recent results by Marshall (2009) and Nie (2012) ...

## Putinar's Positivstellensatz

also holds generically for polynomials of degree $d$ nonnegative on $\mathbf{K}$

# A practical inverse optimization problem

Recall that $y \in \mathbf{K}$ is fixed (given):

**A practical inverse optimization problem $\mathbf{P}_k^{-1}$, $k \in \mathbb{N}$, reads:**

$$\rho_k = \min_{g \in \mathbb{R}[x]_d, \sigma_j} \{\|f - g\| \,:\, g - g(y) = \underbrace{\sigma_0}_{\in \Sigma[\mathbf{x}]_k} + \sum_{j=1}^m g_j \cdot \underbrace{\sigma_j}_{\in \Sigma[\mathbf{x}]_{k-v_j}}$$

- The unknowns, which are the coefficients $(g_\alpha)$ and $(\sigma_{j\alpha})$ of $g \in \mathbb{R}[\mathbf{x}]_d$ and $\sigma_j \in \Sigma[\mathbf{x}]_{k-v_j}$, satisfy a system of LMIs
- The size of the certificate (hence of the LMI's) is controlled by the parameter $k$, the degree of the sos polynomials $\sigma_j$.

# A practical inverse optimization problem

Recall that $y \in \mathbf{K}$ is fixed (given):

**A practical inverse optimization problem $\mathbf{P}_k^{-1}$, $k \in \mathbb{N}$, reads:**

$$\rho_k = \min_{g \in \mathbb{R}[x]_d, \sigma_j} \{\|f - g\| \,:\, g - g(y) = \underbrace{\sigma_0}_{\in \Sigma[\mathbf{x}]_k} + \sum_{j=1}^{m} g_j \cdot \underbrace{\sigma_j}_{\in \Sigma[\mathbf{x}]_{k-v_j}}$$

- The unknowns, which are the coefficients $(g_\alpha)$ and $(\sigma_{j\alpha})$ of $g \in \mathbb{R}[\mathbf{x}]_d$ and $\sigma_j \in \Sigma[\mathbf{x}]_{k-v_j}$, satisfy a system of LMIs
- The size of the certificate (hence of the LMI's) is controlled by the parameter $k$, the degree of the sos polynomials $\sigma_j$.

# A practical inverse optimization problem

Recall that $y \in \mathbf{K}$ is fixed (given):

**A practical inverse optimization problem $\mathbf{P}_k^{-1}$, $k \in \mathbb{N}$, reads:**

$$\rho_k = \min_{g \in \mathbb{R}[x]_d, \sigma_j} \{\|f - g\| \,:\, g - g(y) = \underbrace{\sigma_0}_{\in \Sigma[\mathbf{x}]_k} + \sum_{j=1}^{m} g_j \cdot \underbrace{\sigma_j}_{\in \Sigma[\mathbf{x}]_{k-v_j}}$$

- The unknowns, which are the coefficients $(g_\alpha)$ and $(\sigma_{j\alpha})$ of $g \in \mathbb{R}[\mathbf{x}]_d$ and $\sigma_j \in \Sigma[\mathbf{x}]_{k-v_j}$, satisfy a system of LMIs
- The size of the certificate (hence of the LMI's) is controlled by the parameter $k$, the degree of the sos polynomials $\sigma_j$.

### If the norm $\|h\|$ on $\mathbb{R}[\mathbf{x}]$

is the $\ell_1$, or $\ell_2$, or $\ell_\infty$-norm of the vector of coefficients $(h_\alpha)$ of the polynomial $h$

... then $\mathbf{P}_k^{-1}$ is a semidefinite program

### Theorem

*Let $\mathbf{K} \subset \mathbb{R}^n$ be with nonempty interior. Then for every $2k \geq \deg f$ the practical inverse problem $\mathbf{P}_k^{-1}$ has a optimal solution $g^* \in \mathbb{R}[\mathbf{x}]_d$.*

### If the norm $\|h\|$ on $\mathbb{R}[\mathbf{x}]$

is the $\ell_1$, or $\ell_2$, or $\ell_\infty$-norm of the vector of coefficients $(h_\alpha)$ of the polynomial $h$

... then $\mathbf{P}_k^{-1}$ is a semidefinite program

### Theorem

*Let $\mathbf{K} \subset \mathbb{R}^n$ be with nonempty interior. Then for every $2k \geq \deg f$ the practical inverse problem $\mathbf{P}_k^{-1}$ has a optimal solution $g^* \in \mathbb{R}[\mathbf{x}]_d$.*

# The canonical form of an $\ell_1$-norm solution

Consider the inverse optimization problem $\mathbf{P}_k^{-1}$ with the $\ell_1$-norm.

We consider the case $\mathbf{K}$ compact. With no loss of generality, and up to the change of variable $\mathbf{x}' = \mathbf{x} - \mathbf{y}$ (and possibly after some scaling) one may and will assume that $\mathbf{K} \subseteq [-1, 1]^n$ and $\mathbf{y} \in \mathbf{K}$ is $\mathbf{y} = 0$.

# The canonical form of an $\ell_1$-norm solution

## Theorem

*Let* $\mathbf{K} \subseteq [-1,1]^n$ *be with nonempty interior. Under the* $\ell_1$*-norm, there is an optimal solution* $g^* \in \mathbb{R}[\mathbf{x}]_d$ *of* $\mathbf{P}_k^{-1}$*, with value* $\rho_k$ *and of the form*

$$g^* = f + b'\mathbf{x} + \sum_{i=1}^{n} \lambda_i^* \, x_i^2$$

*for some* $b \in \mathbb{R}^n$ *and nonnegative vector* $\lambda^* \in \mathbb{R}^n$*. And*

$$\rho_k = \|f - g^*\|_1 = \|b\|_1 + \|\lambda^*\|_1 .$$

*Moreover, letting* $J(0) = \{j \,:\, g_j(0) = 0\}$*,*

$$b = -\nabla f(0) + \sum_{j \in J(0)} \gamma_i \nabla g_j(0), \qquad \gamma \geq 0,$$

*for some nonnegative vector* $\gamma$*.*

Observe that in such an optimal solution $g^* \in \mathbb{R}[\mathbf{x}]_d$,

... ONLY $2n$

OUT OF $\binom{n+d}{n}$ $(= O(n^d))$ coefficients of $g^*$ are potentially non zero ... and this ... independently of $d$!

That is, the $\ell_1$-norm criterion INDUCES an optimal solution $g^*$ with a sparse support !!

.... a property already observed in other contexts (e.g. sparse recovery of signals).

Observe that in such an optimal solution $g^* \in \mathbb{R}[\mathbf{x}]_d$,

**... ONLY $2n$**

OUT OF $\binom{n+d}{n}$ $(= O(n^d))$ coefficients of $g^*$ are potentially non zero ... and this ... independently of $d$!

That is, the $\ell_1$-norm criterion INDUCES an optimal solution $g^*$ with a sparse support !!

.... a property already observed in other contexts (e.g. sparse recovery of signals).

# A by-product

As a by product of the inverse optimization problem $\mathbf{P}^{-1}$, we also obtain:

**Theorem**

*Let $f^*$ and $\rho_k$ be the optimal values of $\mathbf{P}$ and $\mathbf{P}_k^{-1}$, respectively, and let $\mathbf{x}^* \in \mathbf{K}$ be an optimal solution of $\mathbf{P}$. Then:*

$$f^* \leq f(y) \leq f^* + \rho_k \cdot \sup_{\alpha \in \mathbb{N}^n_{2d}} |(\mathbf{x}^*)^\alpha|,$$

*and if $\mathbf{K} \subseteq [-1, 1]^n$,*

$$f^* \leq f(y) \leq f^* + \rho_k.$$

And so $\rho_k$ provides an estimate of the how far is $f(y)$ from $f^*$.

# Asymptotics when $k \to \infty$

Recall that $\mathbf{P}^{-1}$ is the ideal inverse problem with value $\rho$.

**Theorem**

*Let $\mathbf{K}$ be with nonempty interior. Let $g_k \in \mathbb{R}[\mathbf{x}]_d$ (resp. $g^* \in \mathbb{R}[\mathbf{x}]_d$) be an optimal solution of $\mathbf{P}_k^{-1}$ (resp. $\mathbf{P}^{-1}$), with associated optimal value $\rho_k$ (resp. $\rho$).*

- *The sequence $(\rho_k)$, $k \in \mathbb{N}$, is monotone nonincreasing and converges to $\hat{\rho} \geq \rho$.*
- *Moreover, every accumulation point $\hat{g} \in \mathbb{R}[\mathbf{x}]_d$ of the sequence $(g_k)$, $k \in \mathbb{N}$, is such that $\hat{g} - \hat{g}(0) \geq 0$ on $\mathbf{K}$ and $\|\hat{g} - f\| = \hat{\rho}$.*
- *Finally, if the polynomial $g^* - g^*(0)$ has a Putinar certificate then $\rho_k = \hat{\rho} = \rho$ for some $k \in \mathbb{N}$.*

# Asymptotics when $k \to \infty$

Recall that $\mathbf{P}^{-1}$ is the ideal inverse problem with value $\rho$.

**Theorem**

*Let* $\mathbf{K}$ *be with nonempty interior. Let* $g_k \in \mathbb{R}[\mathbf{x}]_d$ *(resp.* $g^* \in \mathbb{R}[\mathbf{x}]_d$*) be an optimal solution of* $\mathbf{P}_k^{-1}$ *(resp.* $\mathbf{P}^{-1}$*), with associated optimal value* $\rho_k$ *(resp.* $\rho$*).*

- *The sequence* $(\rho_k)$, $k \in \mathbb{N}$, *is monotone nonincreasing and converges to* $\hat{\rho} \geq \rho$.
- *Moreover, every accumulation point* $\hat{g} \in \mathbb{R}[\mathbf{x}]_d$ *of the sequence* $(g_k)$, $k \in \mathbb{N}$, *is such that* $\hat{g} - \hat{g}(0) \geq 0$ *on* $\mathbf{K}$ *and* $\|\hat{g} - f\| = \hat{\rho}$.
- *Finally, if the polynomial* $g^* - g^*(0)$ *has a Putinar certificate then* $\rho_k = \hat{\rho} = \rho$ *for some* $k \in \mathbb{N}$.

Recall that $\mathbf{P}^{-1}$ is the ideal inverse problem with value $\rho$.

**Theorem**

*Let $\mathbf{K}$ be with nonempty interior. Let $g_k \in \mathbb{R}[\mathbf{x}]_d$ (resp. $g^* \in \mathbb{R}[\mathbf{x}]_d$) be an optimal solution of $\mathbf{P}_k^{-1}$ (resp. $\mathbf{P}^{-1}$), with associated optimal value $\rho_k$ (resp. $\rho$).*

- *The sequence $(\rho_k)$, $k \in \mathbb{N}$, is monotone nonincreasing and converges to $\hat{\rho} \geq \rho$.*
- *Moreover, every accumulation point $\hat{g} \in \mathbb{R}[\mathbf{x}]_d$ of the sequence $(g_k)$, $k \in \mathbb{N}$, is such that $\hat{g} - \hat{g}(0) \geq 0$ on $\mathbf{K}$ and $\|\hat{g} - f\| = \hat{\rho}$.*
- *Finally, if the polynomial $g^* - g^*(0)$ has a Putinar certificate then $\rho_k = \hat{\rho} = \rho$ for some $k \in \mathbb{N}$.*

It has been proved in a number of cases that $f \geq 0$ on $\mathbf{K}$ implies that $f$ has a Putinar certificate, i.e.,

$$f = \underbrace{\sigma_0}_{SOS} + \sum_{j=1}^{m} \underbrace{\sigma_j}_{SOS} g_j,$$

but recent results by Marshall (2006) and Nie (2012) prove that in fact it is a generic property in $\mathbb{R}[\mathbf{x}]_d$!

# $\epsilon$-global minimizer

We would like $\rho_k \to \rho$ (instead of $\rho_k \to \hat{\rho} \geq \rho$) as $k \to \infty$.

**possible ... but need to introduce $\epsilon$-global optimality**

$$\mathbf{P}_{\epsilon}^{-1}: \quad \rho_\epsilon = \min_{g \in \mathbb{R}[\mathbf{x}]_d} \{\|f - g\| \;:\; g(\mathbf{x}) - g(\mathbf{y}) + \epsilon \geq 0, \quad \forall \mathbf{x} \in \mathbf{K}\}$$

and

$$\mathbf{P}_{\epsilon k}^{-1}: \quad \rho_{\epsilon k} = \min_{g \in \mathbb{R}[\mathbf{x}]_d} \{\|f - g\| \;:\; g(\mathbf{x}) - g(\mathbf{y}) + \epsilon = \sigma_0 + \sum_j \sigma_j \, g_j\}$$

with $\deg \sigma_j g_j \leq 2k$ for all $j$.

## Theorem

*Let* $0 < \epsilon_\ell \to 0$ *as* $\ell \to \infty$, *and let* $g_{\ell k} \in \mathbb{R}[\mathbf{x}]_d$ *be an optimal solution of the inverse problem* $\mathbf{P}^{-1}_{\epsilon_\ell k}$.

*For every* $\ell \in \mathbb{N}$ *there exists* $k_\ell$ *such that* $\rho_{\epsilon_\ell k} \leq \rho$ *for all* $k \geq k_\ell$ *and*

$$\rho_{\epsilon_\ell k_\ell} \to \rho \quad \text{and} \quad g_{\ell k_\ell} \to g^* \quad \text{as } \ell \to \infty.$$

# Conclusion

- We have presented a hierarchy of semidefinite programs that provides an approximate solution to inverse polynomial optimization problems.
- For the $\ell_1$-norm criterion, there exists a canonical "sparse" solution.

An interesting issue is to consider problems where the cost function $f$ depends on a parameter $\theta \in \Theta$.

Given $y \in \mathbf{K}$, the inverse problem is now to find a parameter $\theta^* \in \Theta$ that minimizes the error between $f(y, \theta)$ and the optimal value $J(\theta)$ over all $\theta \in \Theta$

**... because in this case**

there might be no parameter value $\theta$ for which $y$ is an optimal solution.

# Conclusion

- We have presented a hierarchy of semidefinite programs that provides an approximate solution to inverse polynomial optimization problems.
- For the $\ell_1$-norm criterion, there exists a canonical "sparse" solution.

An interesting issue is to consider problems where the cost function $f$ depends on a parameter $\theta \in \Theta$.

Given $y \in \mathbf{K}$, the inverse problem is now to find a parameter $\theta^* \in \Theta$ that minimizes the error between $f(y, \theta)$ and the optimal value $J(\theta)$ over all $\theta \in \Theta$

**... because in this case**

there might be no parameter value $\theta$ for which $y$ is an optimal solution.

THANK YOU!