INVERSE POLYNOMIAL OPTIMIZATION

JEAN B. LASSERRE

Abstract. We consider the inverse optimization problem associated with the polynomial program $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ and a given current feasible solution $\mathbf{y} \in \mathbf{K}$. We provide a systematic numerical scheme to compute an inverse optimal solution. That is, we compute a polynomial $\tilde{f}$ (which may be of same degree as f if desired) with the following properties: (a) $\mathbf{y}$ is a global minimizer of $\tilde{f}$ on $\mathbf{K}$ with a Putinar's certificate with an *a priori* degree bound d fixed, and (b), $\tilde{f}$ minimizes $\|f - \tilde{f}\|$ (which can be the ℓ_1, ℓ_2 or ℓ_∞-norm of the coefficients) over all polynomials with such properties. Computing $\tilde{f}_d$ reduces to solving a semidefinite program whose optimal value also provides a bound on how far is $f(\mathbf{y})$ from the unknown optimal value f^*. The size of the semidefinite program can be adapted to to computational capabilities available. Moreover, if one uses the ℓ_1-norm, then $\tilde{f}$ takes a simple and explicit *canonical* form. Some variations are also discussed.

1. Introduction

Let $\mathbf{P}$ be the optimization problem $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$, where

$$\mathbf{K} := \{\mathbf{x} \in \mathbb{R}^n : g_j(\mathbf{x}) \geq 0,\ j = 1, \ldots, m\}, \tag{1.1}$$

for some polynomials $f, (g_j) \subset \mathbb{R}[\mathbf{x}]$. This framework is rather general as it encompasses a large class of important optimization problems, including non convex and discrete optimization problems.

Problem $\mathbf{P}$ is in general NP-hard and one is often satisfied with a local minimum which can be obtained by running some local minimization algorithm among those available in the literature. Typically in such algorithms, at a current iterate (i.e. some feasible solution $\mathbf{y} \in \mathbf{K}$), one checks whether some optimality conditions (e.g. the Karush-Kuhn-Tucker (KKT) conditions) are satisfied within some ϵ-tolerance. However, as already mentioned those conditions are only valid for a local minimum, and in fact, even only for a stationary point of the Lagrangian. Moreover, in many practical situations the criterion f to minimize is subject to modeling errors or is questionable. In such a situation, the practical meaning of a local (or global) minimum f^* (and local (or global) minimizer) also becomes questionable. It could well be that the current solution $\mathbf{y}$ is in fact a global minimizer of an optimization problem $\mathbf{P}'$ with same feasible set as $\mathbf{P}$ but with a different criterion $\tilde{f}$. Therefore, if $\tilde{f}$ is close enough to f, one might not be willing to spend an enormous computing time and effort to find the global (or even local) minimum f^* because one might be already satisfied with the current iterate $\mathbf{y}$ as a global minimizer of $\mathbf{P}'$.

Inverse Optimization is precisely concerned with the above issue of determining a criterion $\tilde{f}$ as close to f as possible, and for which the current solution $\mathbf{y}$ is an

1991 *Mathematics Subject Classification.* 90C26 47J07 65J22 49K30 90C22.

Key words and phrases. Inverse optimization; positivity certificate; mathematical programming; global optimization; semidefinite programming.

optimal solution of $\mathbf{P}'$ with this new criterion $\tilde{f}$. Pioneering work in Control dates back to Freeman and Kokotovic [7] for optimal stabilization. Whereas it was known that every value function of an optimal stabilization problem is also a Lyapunov function for the closed-loop system, in [7] the authors show the converse, that is, every Lyapunov function for every stable closed-loop system is also a *value function* for a meaningful optimal stabilization problem. In optimization, pioneering works in this direction date back to Burton and Toint and [3] for shortest path problems, and Zhang and Liu [17, 18], Huang and Liu [6], and Ahuja and Orlin and [2] for linear programs in the form $\min\{\mathbf{c}'\mathbf{x} \,:\, \mathbf{A}\mathbf{x} \geq \mathbf{b};\, \mathbf{r} \leq \mathbf{x} \leq \mathbf{s}\}$ (and with the ℓ_1-norm). For the latter, the inverse problem is again a linear program of the same form. Similar results also hold for inverse linear programs with the ℓ_∞-norm as shown in Ahuja and Orlin [2] while Zhang et al. [19] provide a column generation method for the inverse shortest path problem. In Heuberger [5] the interested reader will find a nice survey on inverse optimization for linear programming and combinatorial optimization problems. For integer programming, Schaefer [13] characterizes the feasible set of cost vectors $c \in \mathbb{R}^n$ that are candidates for inverse optimality. It is the projection on $\mathbb{R}^n$ of a (lifted) convex polytope obtained from the super-additive dual of integer programs. Unfortunately and as expected, the dimension of of the lifted polyhedron (before projection) is exponential in the input size of the problem. Finally, for linear programs Ahmed and Guan [1] have considered the variant called *inverse optimal value* problem in which one is interested in finding a linear criterion $c \in C \subset \mathbb{R}^n$ for which the optimal value is the closest to a desired specified value. Perhaps surprisingly, they proved that such a problem is NP-hard.

As the reader may immediately guess, in inverse optimization the main difficulty lies in having a tractable characterization of global optimality for a given current point $\mathbf{y} \in \mathbf{K}$ and some candidate criterion $\tilde{f}$. This is why most of all the above cited works address linear programs or combinatorial optimization problems for which some characterization of global optimality is available and can be (sometimes) effectively used for practical computation. For instance, the characterization of global optimality for integer programs described in Schaefer [13] is via the superadditive dual of Wolsey [16, §2] which is exponential in the problem size, and so prevents from its use in practice.

This perhaps explains why inverse (non linear) optimization has not attracted much attention in the past, and it is a pity since inverse optimality could provide an alternative stopping criterion at a feasible solution $\mathbf{y}$ obtained by a (local) optimization algorithm.

The novelty of the present paper is to provide a systematic numerical scheme for computing an inverse optimal solution associated with the polynomial program $\mathbf{P}$ and a given feasible solution $\mathbf{y} \in \mathbf{K}$. It consists of solving a semidefinite program[1] whose size can be adapted to the problem on hand, and so is tractable (at least for moderate size problems and possibly for larger size problems if sparsity is taken into account). Moreover, if one uses the ℓ_1-norm then the inverse-optimal objective function exhibits a simple *canonical* (and sparse) form.

[1] A semidefinite program is a convex (conic) optimization problem that can be solved efficiently. For instance, up to arbitrary (fixed) precision and using some interior point algorithms, it can be solved in time polynomial in the input size of the problem. For more details the interested reader is referred to e.g. Wolkowicz et al. [15] and the many references therein.

Contribution. In this paper we investigate the inverse optimization problem for polynomial optimization problems $\mathbf{P}$ as in (1.1), i.e., in a rather general context which includes nonlinear and nonconvex optimization problems and in particular, 0/1 and mixed integer nonlinear programs. Fortunately, in such a context, Putinar's Positivstellensatz [12] provides us with a very powerful certificate of global optimality that can be adapted to the actual computational capabilities for a given problem size. More precisely, and assuming $\mathbf{y} = 0$ (possibly after a change of variable $\mathbf{x}' = \mathbf{x} - \mathbf{y}$), in the methodology that we propose, one computes the coefficients of a polynomial $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]$ of same degree df as f (or possibly larger degree if desired and/or possibly with some additional constraints), such that:

- 0 is a global optimum of the related problem $\min_{\mathbf{x}}\{\tilde{f}_d(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$, with a Putinar's certificate of optimality with degree bound d (to be explained later).
- $\tilde{f}_d$ minimizes $\|\tilde{f} - f\|_k$ (where depending on k, $\|\cdot\|_k$ is the ℓ_1, ℓ_2 or ℓ_∞-norm of the coefficient vector) over all polynomals $\tilde{f}$ of degree df, having the previous property.

It turns out that the optimal value $\rho_d := \|\tilde{f}_d - f\|_k$ also measures how close is $f(0)$ to the global optimum f^* of $\mathbf{P}$, as we also obtain that $f^* \leq f(0) \leq f^* + \rho_d$ if $k = 1$ and similarly $f^* \leq f(0) \leq f^* + \rho_d \binom{n+df}{n}$ if $k = \infty$.

In addition, for the ℓ_1-norm (and when df is even), we prove that $\tilde{f}_d$ has a *canonical for*, namely

$$\tilde{f}_d \;=\; f + \mathbf{b}'\mathbf{x} + \sum_{i=1}^n (\lambda_i\, x_i^2 + \gamma_i\, x_i^{df}),$$

for some vector $\mathbf{b} \in \mathbb{R}^n$, and nonnegative vectors $\lambda, \gamma \in \mathbb{R}^n$, optimal solutions of a semidefinite program. (For 0/1 problems it further simplifies to $\tilde{f}_d = f + \mathbf{b}'\mathbf{x}$ only.) This canonical form is sparse as $\tilde{f}_d$ differs from f in at most $3n$ entries only. It illustrates the sparsity properties of optimal solutions of ℓ_1-nom minimization problems, already observed in other contexts (e.g., in some compressed sensing applications).

Importantly, to compute $\tilde{f}_d$, one has to solve a semidefinite program of size parametrized by d, where d is chosen so that the size of semidefinite program associated with Putinar's certificate (with degree bound d) is compatible with current semidefinite solvers available. (Of course, even if d is relatively small, one is till restricted to problems of relatively modest size.)

Finally, and when $\mathbf{K}$ is compact, as $d \to \infty$, the sequence of optimal value (ρ_d), $d \in \mathbb{N}$, is shown to converge to the optimal value ρ of the "ideal inverse optimization problem", provided that the optimal solution of the latter satisfies some property.

In addition, one may also consider several additional options:

• Instead of looking for a polynomial $\tilde{f}$ of same degree as f, one might allow polynomials of higher degree, and/or restrict certain coefficients of $\tilde{f}$ to be the same as those of f (e.g. for structural modeling reasons).

• One may restrict $\tilde{f}$ to a certain class of functions, e.g., quadratic polynomials and even convex quadratic polynomials. In the latter important case and if the g_j's that define $\mathbf{K}$ are concave, the procedure to compute an optimal solution $\tilde{f}(\mathbf{x}) = \tilde{\mathbf{b}}'\mathbf{x} + \mathbf{x}'\tilde{\mathbf{Q}}\mathbf{x}$ simplifies and reduces to solving separately a linear program (for computing $\tilde{\mathbf{b}}$) and a semidefinite program (for computing $\tilde{\mathbf{Q}}$).

• One may also define a test of "inverse local optimality" based on the Karush-Kuhn-Tucker optimality conditions, which reduces to solving a linear program instead of a semidefinite program.

The paper is organized as follows. In a first introductory section we present the notation, definitions, and the ideal inverse optimization problem. We then describe how a practical inverse optimization problem reduces to solving a semidefinite program and exhibit the canonical form of the optimal solution for the ℓ_1-norm. We also provide additional results, e.g., an asymptotic analysis when the degree bound in Putinar's certificate increases and also the particular case where one searches for a convex candidate criterion.

2. Notation, definitions and preliminaries

2.1. **Notation and definitions.** Let $\mathbb{R}[\mathbf{x}]$ (resp. $\mathbb{R}[\mathbf{x}]_d$) denote the ring of real polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$ (resp. polynomials of degree at most d), whereas $\Sigma[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_d$) denotes its subset of sums of squares (s.o.s.) polynomials (resp. of s.o.s. of degree at most $2d$). For every $\alpha \in \mathbb{N}^n$ the notation $\mathbf{x}^\alpha$ stands for the monomial $x_1^{\alpha_1} \cdots x_n^{\alpha_n}$ and for every $i \in \mathbb{N}$, let $\mathbb{N}^p_d := \{\beta \in \mathbb{N}^n : \sum_j \beta_j \leq d\}$ whose cardinal is $s(d) = \binom{n+d}{n}$. A polynomial $f \in \mathbb{R}[\mathbf{x}]$ is written

$$\mathbf{x} \mapsto f(\mathbf{x}) \,=\, \sum_{\alpha \in \mathbb{N}^n} f_\alpha \, \mathbf{x}^\alpha,$$

and f can be identified with its vector of coefficients $\mathbf{f} = (f_\alpha)$ in the canonical basis $(\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n$. Denote by $\mathcal{S}^t \subset \mathbb{R}^{t \times t}$ the space of real symmetric matrices, and for any $\mathbf{A} \in \mathcal{S}^t$ the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ is positive semidefinite. For $f \in \mathbb{R}[\mathbf{x}]_d$, let

$$\|f\|_k \,=\, \left\{ \begin{array}{ll} \displaystyle\sum_{\alpha \in \mathbb{N}^n_d} |f_\alpha| & \text{if } k = 1, \\ \displaystyle\sum_{\alpha \in \mathbb{N}^n_d} f_\alpha^2 & \text{if } k = 2, \\ \max\{|f_\alpha| \,:\, \alpha \in \mathbb{N}^n_d\} & \text{if } k = \infty. \end{array} \right.$$

A real sequence $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, has a *representing measure* if there exists some finite Borel measure μ on $\mathbb{R}^n$ such that

$$z_\alpha \,=\, \int_{\mathbb{R}^n} \mathbf{x}^\alpha \, d\mu(\mathbf{x}), \qquad \forall\, \alpha \in \mathbb{N}^n.$$

Given a real sequence $\mathbf{z} = (z_\alpha)$ define the linear functional $L_\mathbf{z} : \mathbb{R}[\mathbf{x}] \to \mathbb{R}$ by:

$$f \;(= \sum_\alpha f_\alpha \mathbf{x}^\alpha) \quad \mapsto L_\mathbf{z}(f) \,=\, \sum_\alpha f_\alpha \, z_\alpha, \qquad f \in \mathbb{R}[\mathbf{x}].$$

Moment matrix. The *moment* matrix associated with a sequence $\mathbf{z} = (z_\alpha)$, $\alpha \in \mathbb{N}^n$, is the real symmetric matrix $\mathbf{M}_d(\mathbf{z})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $z_{\alpha+\beta}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. Alternatively, let $\mathbf{v}_d(\mathbf{x}) \in \mathbb{R}^{s(d)}$ be the vector $(\mathbf{x}^\alpha)$, $\alpha \in \mathbb{N}^n_d$, and define the matrices $(\mathbf{B}_\alpha) \subset \mathcal{S}^{s(d)}$ by

$$\mathbf{v}_d(\mathbf{x}) \, \mathbf{v}_d(\mathbf{x})^T \,=\, \sum_{\alpha \in \mathbb{N}^n_{2d}} \mathbf{B}_\alpha \, \mathbf{x}^\alpha, \qquad \forall \mathbf{x} \in \mathbb{R}^n. \tag{2.1}$$

Then $\mathbf{M}_d(\mathbf{z}) = \sum_{\alpha \in \mathbb{N}^n_{2d}} z_\alpha \, \mathbf{B}_\alpha$.

If $\mathbf{z}$ has a representing measure μ then $\mathbf{M}_d(\mathbf{z}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(\mathbf{z})\mathbf{f}\rangle \,=\, \int f^2\, d\mu \,\geq 0, \qquad \forall\, \mathbf{f} \,\in \mathbb{R}^{s(d)}.$$

Localizing matrix. With $\mathbf{z}$ as above and $g \in \mathbb{R}[\mathbf{x}]$ (with $g(\mathbf{x}) = \sum_\gamma g_\gamma \mathbf{x}^\gamma$), the *localizing* matrix associated with $\mathbf{z}$ and g is the real symmetric matrix $\mathbf{M}_d(g\,\mathbf{z})$ with rows and columns indexed by $\mathbb{N}^n_d$, and whose entry (α, β) is just $\sum_\gamma g_\gamma z_{\alpha+\beta+\gamma}$, for every $\alpha, \beta \in \mathbb{N}^n_d$. Alternatively, let $\mathbf{C}_\alpha \in \mathcal{S}^{s(d)}$ be defined by:

$$g(\mathbf{x})\, \mathbf{v}_d(\mathbf{x})\, \mathbf{v}_d(\mathbf{x})^T \,=\, \sum_{\alpha \in \mathbb{N}^n_{2d+\deg g}} \mathbf{C}_\alpha\, \mathbf{x}^\alpha, \qquad \forall \mathbf{x} \in \mathbb{R}^n. \tag{2.2}$$

Then $\mathbf{M}_d(g\,\mathbf{z}) = \sum_{\alpha \in \mathbb{N}^n_{2d+\deg g}} z_\alpha\, \mathbf{C}_\alpha$.

If $\mathbf{z}$ has a representing measure μ whose support is contained in the set $\{\mathbf{x} : g(\mathbf{x}) \geq 0\}$ then $\mathbf{M}_d(g\,\mathbf{z}) \succeq 0$ because

$$\langle \mathbf{f}, \mathbf{M}_d(g\,\mathbf{z})\mathbf{f}\rangle \,=\, \int f^2\, g\, d\mu \,\geq 0, \qquad \forall\, \mathbf{f} \,\in \mathbb{R}^{s(d)}.$$

With $\mathbf{K}$ as in (1.1), let $g_0 \in \mathbb{R}[\mathbf{x}]$ be the constant polynomial $\mathbf{x} \mapsto g_0(\mathbf{x}) = 1$, and for every $j = 0, 1, \ldots, m$, let $v_j := \lceil (\deg g_j)/2 \rceil$.

Definition 2.1. *With $d, k \in \mathbb{N}$ and $\mathbf{K}$ as in (1.1), let $Q_k(g) \subset \mathbb{R}[\mathbf{x}]$ and $Q^d_k \subset \mathbb{R}[\mathbf{x}]_d$ be the convex cones:*

$$Q(g) \;:=\; \left\{ \sum_{k=0}^m \sigma_j\, g_j \;:\; \sigma_j \in \Sigma[\mathbf{x}] \quad j = 1, \ldots, m \right\}. \tag{2.3}$$

$$Q_k(g) \;:=\; \left\{ \sum_{k=0}^m \sigma_j\, g_j \;:\; \sigma_j \in \Sigma[\mathbf{x}]_{k-v_j}, \quad j = 1, \ldots, m \right\}. \tag{2.4}$$

$$Q^d_k(g) \;:=\; Q_k(g) \,\cap\, \mathbb{R}[\mathbf{x}]_d \tag{2.5}$$

We say that every element $h \in Q_k(g)$ has a Putinar's certificate of nonnegativity on $\mathbf{K}$, with degree bound k.

The cone $Q(g)$ is called the quadratic module associated with the g_j's. Obviously, if $h \in Q(g)$ the associated s.o.s. polynomials σ_j's provide a certificate of nonnegativity of h on $\mathbf{K}$. The cone $Q(g)$ is said to be *Archimedean* if and only if

$$\mathbf{x} \mapsto M - \|\mathbf{x}\|^2 \,\in\, Q(g) \quad \text{for some } M > 0. \tag{2.6}$$

Let $\mathrm{Psd}_d(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_d$ be the convex cone of polynomials of degree at most d, nonnegative on $\mathbf{K}$. The name "Putinar's certificate" is coming from the following Putinar's Positivstellensatz.

Theorem 2.2 (Putinar's Positivstellensatz [12])**.** *Let $\mathbf{K}$ be as in (1.1) and assume that $Q(g)$ is Archimedean. Then every polynomial $f \in \mathbb{R}[\mathbf{x}]$ strictly positive on $\mathbf{K}$ belongs to $Q(g)$. In addition,*

$$\mathrm{cl}\left(\bigcup_{k=0}^{\infty} Q^d_k(g) \right) \,=\, \mathrm{Psd}_d(\mathbf{K}), \qquad \forall d \in \mathbb{N}. \tag{2.7}$$

The first statement is just Putinar's Positivstellensatz [12] whereas the second statement is an easy consequence. Indeed let $f \in \mathrm{Psd}_d(\mathbf{K})$. If $f > 0$ on $\mathbf{K}$ then $f \in Q_k^d(g)$ for some k. If $f(\mathbf{x}) = 0$ for some $\mathbf{x} \in \mathbf{K}$, let $f_n := f + 1/n$, so that $f_n > 0$ on $\mathbf{K}$ for every $n \in \mathbb{N}$. But then $f_n \in \cup_{k=0}^{\infty} Q_k^d(g)$ and the result follows because $\|f_n - f\|_1 \to 0$ as $n \to \infty$.

So Putinar's Positivstellensatz is what we need to *certify* that a polynomial is nonnegative on $\mathbf{K}$, and in particular the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - f(\mathbf{y})$ for the inverse optimization problem associated with a feasible solution $\mathbf{y} \in \mathbf{K}$.

2.2. **The ideal inverse problem.** Let $\mathbf{P}$ be the global optimization problem $f^* = \min_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ with $\mathbf{K} \subset \mathbb{R}^n$ as in (1.1), and $f \in \mathbb{R}[\mathbf{x}]_{df}$ where $df := \deg f$.

Identifying a polynomial $f \in \mathbb{R}[\mathbf{x}]_{df}$ with its vector of coefficients $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(df)}$, one may and will identify $\mathbb{R}[\mathbf{x}]_{df}$ with the vector space $\mathbb{R}^{s(df)}$, i.e., $\mathbb{R}[\mathbf{x}]_{df} \ni f \leftrightarrow \mathbf{f} \in \mathbb{R}^{s(df)}$. Similarly, the convex cone $\mathrm{Psd}_{df}(\mathbf{K}) \subset \mathbb{R}[\mathbf{x}]_{df}$ can be identified with the convex cone $\{\mathbf{h} \in \mathbb{R}^{s(df)} : \mathbf{h} \leftrightarrow h \in \mathrm{Psd}_{df}(\mathbf{K})\}$ of $\mathbb{R}^{s(df)}$. So in the sequel, and unless if necessary, we will not distinguish between f and $\mathbf{f}$.

Next, let $\mathbf{y} \in \mathbf{K}$ and $k \in \{1, 2, \infty\}$ both fixed, and consider the following optimization problem $\mathcal{P}$

$$\mathcal{P}: \quad \rho^k = \min_{\tilde{f} \in \mathbb{R}[\mathbf{x}]_{df}} \{ \|\tilde{f} - f\|_k \,:\, \mathbf{x} \mapsto \tilde{f}(\mathbf{x}) - \tilde{f}(\mathbf{y}) \,\in\, \mathrm{Psd}_{df}(\mathbf{K}) \}. \tag{2.8}$$

Theorem 2.3. *Problem (2.8) has an optimal solution $\tilde{f}^* \in \mathbb{R}[\mathbf{x}]_{df}$. In addition, $\rho^k = 0$ if and only if $\mathbf{y}$ is an optimal solution of $\mathbf{P}$.*

Proof. Obviously the constant polynomial $\mathbf{x} \mapsto \tilde{f}(\mathbf{x}) := 1$ is a feasible solution with associated value $\delta := \|\tilde{f} - f\|_k$. Moreover the optimal value of (2.8) is bounded below by 0. Consider a minimizing sequence $(\tilde{f}^j) \subset \mathbb{R}[\mathbf{x}]_{df}$, $j \in \mathbb{N}$, hence such that $\|\tilde{f}^j - f\|_k \to \rho^k$ as $j \to \infty$. As we have $\|\tilde{f}^j - f\|_k \leq \delta$ for every j, the sequence $(\tilde{f}^j)$ belongs to the ℓ_k-ball $\mathbf{B}_k(f) := \{\tilde{f} \in \mathbb{R}[\mathbf{x}]_{df} : \|\tilde{f} - f\|_k \leq \delta\}$, a compact set. Therefore, there is an element $\tilde{f}^* \in \mathbf{B}_k(f)$ and a subsequence (j_t), $t \in \mathbb{N}$, such that $\tilde{f}^{j_t} \to \tilde{f}^*$ as $t \to \infty$. Let $\mathbf{x} \in \mathbf{K}$ be fixed arbitrary. Obviously $(0 \leq)\, \tilde{f}^{j_t}(\mathbf{x}) - \tilde{f}^{j_t}(\mathbf{y}) \to \tilde{f}^*(\mathbf{x}) - \tilde{f}^*(\mathbf{y})$ as $t \to \infty$, which implies $\tilde{f}^*(\mathbf{x}) - \tilde{f}^*(\mathbf{y}) \geq 0$, and so, as $\mathbf{x} \in \mathbf{K}$ was arbitrary, $\tilde{f}^* - \tilde{f}^*(\mathbf{y}) \geq 0$ on $\mathbf{K}$, i.e., $\tilde{f}^* - \tilde{f}^*(\mathbf{y}) \in \mathrm{Psd}_{df}(\mathbf{K})$. Finally, we also obtain the desired result

$$\rho^k \,=\, \lim_{j\to\infty} \|\tilde{f}^j - f\|_k \,=\, \lim_{t\to\infty} \|\tilde{f}^{j_t} - f\|_k \,=\, \|\tilde{f}^* - f\|_k.$$

Next, if $\mathbf{y}$ is an optimal solution of $\mathbf{P}$ then $\tilde{f} := f$ is an optimal solution of $\mathcal{P}$ with value $\rho^k = 0$. Conversely, if $\rho^k = 0$ then $\tilde{f}^* = f$, and so by feasibility of $\tilde{f}^*\,(= f)$ for (2.8), $f(\mathbf{x}) \geq f(\mathbf{y})$ for all $\mathbf{x} \in \mathbf{K}$, which shows that $\mathbf{y}$ is an optimal solution of $\mathbf{P}$. □

Theorem 2.3 states that the ideal inverse optimization problem is well-defined. However, even though $\mathrm{Psd}_{df}(\mathbf{K})$ is a finite dimensional convex cone, it has no simple and tractable characterization to be used for practical computation. Therefore one needs an alternative and more tractable version of problem $\mathcal{P}$. Fortunately, we next show that in the polynomial context such a formulation exists, thanks to the powerful Putinar's Positivstellensatz (Theorem 2.2 above).

3. Main result

As the ideal inverse problem is intractable, we here provide tractable formulations whose size depends on a parameter $d \in \mathbb{N}$. If the polynomial $\tilde{f}^*$ in Theorem 2.3 belongs to $Q(g)$ then when d increases the associated optimal value ρ_d^k converges in finitely many steps to the optimal value ρ^k of the ideal problem (2.8), and $\tilde{f}^*$ can be obtained by solving finitely many semidefinite programs. With no loss of generality, i.e., up to some change of variable $\mathbf{x}' = \mathbf{x} - \mathbf{y}$, we may and will assume that $\mathbf{y} = 0 \in \mathbf{K}$.

3.1. **A practical inverse problem.** With $d \in \mathbb{N}$ fixed, consider the following optimization problem $\mathbf{P}_d$:

$$
(3.1)\qquad
\begin{aligned}
\mathbf{P}_d: \quad \rho_d^k := \min_{\tilde{f}, \sigma_j \in \mathbb{R}[\mathbf{x}]} \quad & \|f - \tilde{f}\|_k \\
\text{s.t.} \quad & \tilde{f}(\mathbf{x}) - \tilde{f}(0) \;=\; \sum_{j=0}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n \\
& \tilde{f} \in \mathbb{R}[\mathbf{x}]_{df};\ \sigma_j \in \Sigma[\mathbf{x}]_{d - v_j}, \quad j = 0, 1, \ldots, m,
\end{aligned}
$$

where $df = \deg f$, and $v_j = \lceil (\deg g_j)/2 \rceil$, $j = 1, \ldots, m$.

The parameter $d \in \mathbb{N}$ impacts (3.1) by the maximum degree allowed for the s.o.s. weights $(\sigma_j) \subset \Sigma[\mathbf{x}]$ in Putinar's certificate for the polynomial $\mathbf{x} \mapsto \tilde{f}(\mathbf{x}) - \tilde{f}(0)$, and so the higher d the lower ρ_d^k. Next, observe that in (3.1), the constraint

$$
\tilde{f}(\mathbf{x}) - \tilde{f}(0) \;=\; \sum_{j=0}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}), \quad \forall \mathbf{x} \in \mathbb{R}^n,
$$

is equivalent to stating that $\tilde{f}(\mathbf{x}) - \tilde{f}(0) \in Q_d^{df}(g)$, with $Q_d^{df}(g)$ as in (2.5). Therefore, in particular, $\tilde{f}(\mathbf{x}) \geq \tilde{f}(0)$ for all $\mathbf{x} \in \mathbf{K}$, and so 0 is a global minimum of $\tilde{f}$ on $\mathbf{K}$. So $\mathbf{P}_d$ is a strenghtening of $\mathcal{P}$ in that one has replaced the constraint $\tilde{f} - \tilde{f}(0) \in \mathrm{Psd}_{df}(\mathbf{K})$ with the stronger condition $\tilde{f} - \tilde{f}(0) \in Q_d^{df}(g)$. And so $\rho^k \leq \rho_d^k$ for all $d \in \mathbb{N}$. However, as we next see, $\mathbf{P}_d$ is a tractable optimization problem with nice properties. Indeed, $\mathbf{P}_d$ is a convex optimization problem and even a semidefinite program. For instance, if $k = 1$ one may rewrite $\mathbf{P}_d$ as:

$$
(3.2)\qquad
\begin{aligned}
\rho_d^1 := \min_{\tilde{f}, \lambda_\alpha, \mathbf{Z}_j} \quad & \sum_{\alpha \in \mathbb{N}^n_{df} \setminus \{0\}} \lambda_\alpha \\
\text{s.t.} \quad & \lambda_\alpha + \tilde{f}_\alpha \geq f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \lambda_\alpha - \tilde{f}_\alpha \geq -f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \langle \mathbf{Z}_0, \mathbf{B}_\alpha \rangle + \sum_{j=1}^{m} \langle \mathbf{Z}_j, \mathbf{C}_\alpha^j \rangle \;=\; \begin{cases} \tilde{f}_\alpha, & \text{if } 0 < |\alpha| \leq df \\ 0, & \text{if } \alpha = 0 \text{ or } |\alpha| > df \end{cases} \\
& \mathbf{Z}_j \succeq 0, \quad j = 0, 1, \ldots, m.
\end{aligned}
$$

with $\mathbf{B}_\alpha$ as in (2.1) and $\mathbf{C}_\alpha^j$ as in (2.2) (with g_j in lieu of g). If $k = \infty$ then one may rewrite $\mathbf{P}_d$ as:

$$
(3.3)\qquad \begin{array}{rl}
\rho_d^\infty := \displaystyle\min_{\tilde f,\lambda,\mathbf{Z}_j} & \lambda \\
\text{s.t.} & \lambda + \tilde f_\alpha \geq f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \lambda - \tilde f_\alpha \geq -f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \langle \mathbf{Z}_0, \mathbf{B}_\alpha\rangle + \displaystyle\sum_{j=1}^m \langle \mathbf{Z}_j, \mathbf{C}^j_\alpha\rangle \;=\; \left\{\begin{array}{l} \tilde f_\alpha, \;\text{ if } 0 < \vert\alpha\vert \leq \, df \\ 0, \;\text{ if } \alpha = 0 \text{ or } \vert\alpha\vert > df \end{array}\right. \\
& \mathbf{Z}_j \succeq 0, \quad j = 0, 1, \ldots, m,
\end{array}
$$

and finally, if $k = 2$ then one may rewrite $\mathbf{P}_d$ as:

$$
(3.4)\qquad \begin{array}{rl}
\rho_d^2 := \displaystyle\min_{\tilde f,\lambda,\mathbf{Z}_j} & \displaystyle\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} \lambda_\alpha \\
\text{s.t.} & \left[\begin{array}{cc} \lambda_\alpha & \tilde f_\alpha - f_\alpha \\ \tilde f_\alpha - f_\alpha & 1 \end{array}\right] \succeq 0, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \langle \mathbf{Z}_0, \mathbf{B}_\alpha\rangle + \displaystyle\sum_{j=1}^m \langle \mathbf{Z}_j, \mathbf{C}^j_\alpha\rangle \;=\; \left\{\begin{array}{l} \tilde f_\alpha, \;\text{ if } 0 < \vert\alpha\vert \leq \, df \\ 0, \;\text{ if } \alpha = 0 \text{ or } \vert\alpha\vert > df \end{array}\right. \\
& \mathbf{Z}_j \succeq 0, \quad j = 0, 1, \ldots, m.
\end{array}
$$

Remark 3.1. Observe that in any feasible solution $(\tilde f, \lambda, \mathbf{Z}_j)$ in all formulations (3.2)-(3.4), $\tilde f_0$ plays no role in the constraints of (3.1), but since we minimize $\Vert\tilde f - f\Vert_k$ then it is always optimal to set $\tilde f_0 = f_0$. That is, $\tilde f_d(0) = \tilde f_0 = f_0 = f(0)$.

Sparsity. The semidefinite program (3.1)-(3.2) has $m + 1$ Linear Matrix Inequalities (LMI's) $\mathbf{Z}_j \succeq 0$ of size $O(n^d)$, which limits its application to small to medium size problems $\mathbf{P}$. However large scale problems usually exhibit sparsity patterns which sometimes can be exploited. For instance, in [10] we have provided a specialized "sparse" version of Theorem 2.2 for problems with structured sparsity as described in Waki et al. [14]. Hence, with this specialized version of Putinar's Positivstellensatz, one obtains a *sparse* positivity certificate which when substituted in (3.1), would permit to solve (3.1) for problems of much larger size. Typically, in [14] the authors have applied the "sparse semidefinite relaxations" to problem $\mathbf{P}$ with up to 1000 variables! Moreover, the *running intersection property* that must satisfy the sparsity pattern for convergence guarantee of such relaxations [10], is *not* needed in the present context of inverse optimization. This is because one imposes $\tilde f$ to satisfy this specialized Putinar's Positivstellensatz.

3.2. **Duality.** The semidefinite program dual of (3.2) reads

$$
(3.5)\qquad \left\{\begin{array}{ll}
\displaystyle\max_{\mathbf{u},\mathbf{v}\geq 0,\mathbf{z}} & \displaystyle\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} f_\alpha(u_\alpha - v_\alpha)\; (= L_{\mathbf{z}}(f(0) - f)) \\
\text{s.t.} & u_\alpha + v_\alpha \;\leq\; 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& u_\alpha - v_\alpha + z_\alpha \;=\; 0, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\} \\
& \mathbf{M}_d(\mathbf{z}),\, \mathbf{M}_{d-v_j}(g_j\,\mathbf{z}) \;\succeq\; 0, \quad j = 1, \ldots, m,
\end{array}\right.
$$

while the semidefinite program dual of (3.3) reads

$$
(3.6)\qquad \left\{\begin{array}{ll}
\displaystyle\max_{\mathbf{u},\mathbf{v}\geq 0,\mathbf{z}} & \displaystyle\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} f_\alpha(u_\alpha - v_\alpha)\ (= L_{\mathbf{z}}(f(0)-f))\\
\text{s.t.} & \displaystyle\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} u_\alpha + v_\alpha \ \leq\ 1\\
& u_\alpha - v_\alpha + z_\alpha \ =\ 0,\quad \forall\alpha\in\mathbb{N}^n_{df}\setminus\{0\}\\
& \mathbf{M}_d(\mathbf{z}),\, \mathbf{M}_{d-v_j}(g_j\,\mathbf{z})\ \succeq\ 0,\quad j=1,\ldots,m,
\end{array}\right.
$$

and the semidefinite program dual of (3.4) reads

$$
(3.7)\qquad \left\{\begin{array}{ll}
\displaystyle\max_{\mathbf{z},\Delta_\alpha} & \displaystyle\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} \left\langle \Delta_\alpha, \left[\begin{array}{cc} 0 & f_\alpha\\ f_\alpha & -1\end{array}\right]\right\rangle\\
\text{s.t.} & \left\langle \Delta_\alpha, \left[\begin{array}{cc} 1 & 0\\ 0 & 0\end{array}\right]\right\rangle \leq 1,\quad \forall\alpha\in\mathbb{N}^n_{df}\setminus\{0\}\\
& \left\langle \Delta_\alpha, \left[\begin{array}{cc} 0 & 1\\ 1 & 0\end{array}\right]\right\rangle + z_\alpha \ =\ 0,\quad \forall\alpha\in\mathbb{N}^n_{df}\setminus\{0\}\\
& \mathbf{M}_d(\mathbf{z}),\, \mathbf{M}_{d-v_j}(g_j\,\mathbf{z})\ \succeq\ 0,\quad j=1,\ldots,m\\
& \Delta_\alpha\succeq 0,\quad \forall\alpha\in\mathbb{N}^n_{df}\setminus\{0\}
\end{array}\right.
$$

One may show that one may replace the criterion in (3.7) with the equivalent criterion

$$
\max_{\mathbf{z}}\left\{L_{\mathbf{z}}(f(0)-f)) - \frac{1}{4}\sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}}\left(\int_{\mathbf{K}}\mathbf{x}^\alpha\,d\mu(\mathbf{x})\right)^2\right\}.
$$

Lemma 3.2. *Assume that $\mathbf{K}\subset\mathbb{R}^n$ has nonempty interior. Then there is no duality gap between the semidefinite programs (3.2) and (3.5), (3.3) and (3.6), and (3.4) and (3.7). Moreover, all semidefinite programs (3.2), (3.3) and (3.4) have an optimal solution $\tilde f_d\in\mathbb{R}[\mathbf{x}]_{df}$.*

Proof. The proof is detailed for the case $k=1$ and omitted for the cases $k=2$ and $k=\infty$ because it is very similar. Observe that $\rho^1_d\geq 0$ and the constant polynomial $\tilde f(\mathbf{x})=0$ for all $\mathbf{x}\in\mathbb{R}^n$, is an obviously feasible solution of (3.1) (hence of (3.2)). Therefore ρ^1_d being finite, it suffices to prove that Slater's condition[2] holds for the dual (3.5). Then the conclusion of Lemma 3.2 follows from a standard result of convex optimization. Let μ be the finite Borel measure defined by

$$
\mu(B) := \int_{B\cap\mathbf{K}} \mathrm{e}^{-\|\mathbf{x}\|^2}d\mathbf{x},\qquad \forall B\in\mathcal{B}
$$

(where $\mathcal{B}$ is the usual Borel σ-field), and let $\mathbf{z}=(z_\alpha)$, $\alpha\in\mathbb{N}^n_{2d}$, with

$$
z_\alpha := \kappa\int_{\mathbf{K}}\mathbf{x}^\alpha\,d\mu(\mathbf{x}),\qquad \alpha\in\mathbb{N}^n_{2d},
$$

for some $\kappa>0$ sufficiently small to ensure that

$$
(3.8)\qquad \kappa\left|\int\mathbf{x}^\alpha\,d\mu(\mathbf{x})\right| < 1,\qquad \forall\alpha\in\mathbb{N}^n_{2d}\setminus\{0\}.
$$

[2]Slater's condition holds if there exists a strictly feasible solution, and so for the dual (3.5), if there exists $\mathbf{z}$ such that $\mathbf{M}_d(\mathbf{z}),\mathbf{M}_{d-v_j}(g_j\,\mathbf{z})\succ 0$, $j=1,\ldots,m$, and $u_\alpha+v_\alpha<1$, $\forall\alpha\in\mathbb{N}^n_{df}\setminus\{0\}$. Then from a standard result in convex optimization, there is no duality gap between (3.2) and (3.5), and if the values are bounded then (3.2) has an optimal solution.

Define $u_\alpha = \max[0, -z_\alpha]$ and $v_\alpha = \max[0, z_\alpha]$, $\alpha \in \mathbb{N}^n_{df} \setminus \{0\}$, so that $u_\alpha + v_\alpha < 1$, $\alpha \in \mathbb{N}^n_{2d} \setminus \{0\}$. Hence $(u_\alpha, v_\alpha, \mathbf{z})$ is a feasible solution of (3.5). In addition, $\mathbf{M}_d(\mathbf{z}) \succ 0$ and $\mathbf{M}_{d-v_j}(g_j\,\mathbf{z}) \succ 0$, $j = 1, \ldots, m$, because $\mathbf{K}$ has nonempty interior, and so Slater's condition holds for (3.5), the desired result.

If $k = \infty$ one chooses $\mathbf{z}$ such that

$$\kappa \sum_{\alpha \in \mathbb{N}^n_{df} \setminus \{0\}} \left| \int \mathbf{x}^\alpha \, d\mu(\mathbf{x}) \right| < 1,$$

and if $k = 2$ then one chooses $\mathbf{z}$ as in (3.8) and $\Delta_\alpha := \left[\begin{array}{cc} 1/2 & \kappa_\alpha \\ \kappa_\alpha & 1 \end{array} \right] \succ 0$, for all $\alpha \in \mathbb{N}^n_{df} \setminus \{0\}$, such that

$$2\kappa_\alpha := \kappa \int \mathbf{x}^\alpha \, d\mu(\mathbf{x}), \qquad \forall \alpha \in \mathbb{N}^n_{2d} \setminus \{0\}.$$

$\square$

Theorem 3.3. *Assume that* $\mathbf{K}$ *in (1.1) has nonempty interior, and let* $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer of* $\mathbf{P}$ *with optimal value* f^**, and let* $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]_{df}$ *be an optimal solution of* $\mathbf{P}_d$ *in (3.1) with optimal value* ρ^k_d*. Then:*

(a) $0 \in \mathbf{K}$ *is a global minimizer of the problem* $\tilde{f}^*_d = \min_{\mathbf{x}} \{\tilde{f}_d(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$*. In particular, if* $\rho^k_d = 0$ *then* $\tilde{f}_d = f$ *and* 0 *is a global minimizer of* $\mathbf{P}$*.*

(b) *If* $k = 1$ *then* $f^* \le f(0) \le f^* + \rho^1_d \sup_{\alpha \in \mathbb{N}^n_{df}} |(\mathbf{x}^*)^\alpha|$*. In particular, if* $\mathbf{K} \subseteq [-1, 1]^n$ *then* $f^* \le f(0) \le f^* + \rho^1_d$*.*

(c) *If* $k = \infty$ *then* $f^* \le f(0) \le f^* + \rho^\infty_d \sum_{\alpha \in \mathbb{N}^n_{df}} |(\mathbf{x}^*)^\alpha|$*. In particular if* $\mathbf{K} \subseteq [-1, 1]^n$ *then* $f^* \le f(0) \le f^* + s(df)\, \rho^\infty_d$*.*

Proof. (a) Existence of $\tilde{f}_d$ is guaranteed by Lemma 3.2. From the constraints of (3.1) we have: $\tilde{f}_d(\mathbf{x}) - \tilde{f}(0) = \sum_{j=0}^m \sigma_j(\mathbf{x})\, g_j(\mathbf{x})$ which implies that $\tilde{f}_d(\mathbf{x}) \ge \tilde{f}(0)$ for all $\mathbf{x} \in \mathbf{K}$, and so 0 is a global minimizer of the optimization problem $\mathbf{P}'$: $\min_{\mathbf{x}} \{\tilde{f}_d(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K}\}$.

(b) Let $\mathbf{x}^* \in \mathbf{K}$ be a global minimizer of $\mathbf{P}$. Observe that with $k = 1$,

$$\begin{aligned} f^* = f(\mathbf{x}^*) &= \underbrace{f(\mathbf{x}^*) - \tilde{f}_d(\mathbf{x}^*)} + \underbrace{\tilde{f}_d(\mathbf{x}^*) - \tilde{f}_d(0)}_{\ge 0} + \tilde{f}_d(0) \\ &\ge \tilde{f}_d(0) - |\tilde{f}_d(\mathbf{x}^*) - f(\mathbf{x}^*)| \\ &\ge \tilde{f}_d(0) - \|\tilde{f}_d - f\|_1 \times \sup_{\alpha \in \mathbb{N}^n_{df}} |(\mathbf{x}^*)^\alpha| \\ &\ge f(0) - \rho^1_d \sup_{\alpha \in \mathbb{N}^n_{df}} |(\mathbf{x}^*)^\alpha| \end{aligned} \tag{3.9}$$

since $\tilde{f}_d(0) = f(0)$; see Remark 3.1.

(c) The proof is similar to that of (b) using that with $k = \infty$,

$$|\tilde{f}_d(\mathbf{x}) - f(\mathbf{x})| \ge \left(\sup_{\alpha \in \mathbb{N}^n_{df}} |\tilde{f}_{d\alpha} - f_\alpha| \right) \times \sum_{\alpha \in \mathbb{N}^n_{df}} |\mathbf{x}^\alpha|.$$

□

So not only Theorem 3.3 states that 0 is the global optimum of the optimization problem $\min\{\tilde{f}_d(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$, but it also states that the optimal value ρ_d^k also measures how far is $f(0)$ from the optimal value f^* of the initial problem $\mathbf{P}$. Moreover, observe that Theorem 3.3 merely requires existence of a minimizer and nonemptyness of $\mathbf{K}$. In particular, $\mathbf{K}$ may not be compact.

3.3. **A canonical form for the ℓ_1-norm.** If one uses the ℓ_1-norm when df is even, then the optimal solution $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]_{df}$ in Theorem 3.3 (with $k = 1$) takes a particularly simple *canonical* form:

Theorem 3.4. *Assume that $\mathbf{K}$ has a nonempty interior and let $\mathbf{x}^* \in \mathbf{K}$ be a global minimizer of $\mathbf{P}$ with optimal value f^*. Let $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]_{df}$ be an optimal solution of $\mathbf{P}_d$ in (3.1) with optimal value ρ_d^1 for the ℓ_1-norm. Then $\tilde{f}_d$ is of the form:*

$$\tilde{f}_d(\mathbf{x}) \,=\, f(\mathbf{x}) + \mathbf{b}'\mathbf{x} + \sum_{i=1}^n \left(\lambda_i^*\, x_i^2 + \gamma_i^*\, x_i^{df}\right), \tag{3.10}$$

for some vector $\mathbf{b} \in \mathbb{R}^n$ and some nonnegative vectors $\lambda^*, \gamma^* \in \mathbb{R}^n$, optimal solution of the semidefinite program:

$$\begin{array}{rl} \rho_d^1 := \displaystyle\min_{\lambda,\gamma,\mathbf{b}} & \|\mathbf{b}\|_1 + \displaystyle\sum_{i=1}^n (\lambda_i + \gamma_i) \\ s.t. & f - f(0) + \mathbf{b}'\mathbf{x} + \displaystyle\sum_{i=1^n} \left(\lambda_i\, x_i^2 + \gamma_i\, x_i^{df}\right) \in P_{2d}(g), \quad \lambda, \gamma \geq 0. \end{array} \tag{3.11}$$

Proof. Notice that the dual (3.5) of (3.2) is equivalent to:

$$\left\{ \begin{array}{ll} \displaystyle\max_{\mathbf{z}} & L_{\mathbf{z}}(f(0) - f)) \\ \text{s.t.} & \mathbf{M}_d(\mathbf{z}),\, \mathbf{M}_{d-v_j}(g_j\,\mathbf{z}) \succeq 0, \quad j = 1, \ldots, m, \\ & |z_\alpha| \leq 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \{0\}. \end{array} \right. \tag{3.12}$$

But $\mathbf{M}_d(\mathbf{z}) \succeq 0 \Rightarrow \mathbf{M}_{df/2}(\mathbf{z}) \succeq 0$. On the other hand, recalling that df is even and $\mathbf{M}_{df/2}(\mathbf{z}) \succeq 0$, one may use same arguments as those used in Lasserre and Netzer [11, Lemma 4.1, 4.2], to obtain $|z_\alpha| \leq \max_i\{\max[L_{\mathbf{z}}(x_i^2), L_{\mathbf{x}}(x_i^{df})]\}$, for very $\alpha \in \mathbb{N}^n_{df}$ with $1 < |\alpha| \leq df$. Hence in (3.12) one may replace the constraint $|z_\alpha| \leq 1$ for all $\alpha \in \mathbb{N}^n_{df} \setminus \{0\}$ with the $3n$ inequality constraints:

$$\pm L_{\mathbf{z}}(x_i) \leq 1,\; L_{\mathbf{z}}(x_i^2) \leq 1,\; L_{\mathbf{x}}(x_i^{df}) \leq 1,\; i = 1, \ldots, n. \tag{3.13}$$

And so the dual of the modified SDP is now

$$\left\{ \begin{array}{ll} \displaystyle\max_{\mathbf{b}^1,\mathbf{b}^2,\lambda,\gamma} & \displaystyle\sum_{i=1}^n \left((b_i^1 + b_i^2) + \lambda_i + \gamma_i\right) \\ \text{s.t.} & f - f(0) + (\mathbf{b}^1 - \mathbf{b}^2)'\mathbf{x} + \sum_{i=1}^n \left(\lambda_i x_i^2 + \gamma_i\, x_i^{df}\right) \,\in\, P_{2d}(g) \\ & \mathbf{b}^1, \mathbf{b}^2, \lambda, \gamma \geq 0 \end{array} \right.$$

which is equivalent to (3.11). □

From the proof of Theorem 3.4, this special form of $\tilde{f}_d$ is specific to the ℓ_1-norm, which yields the constraint $|z_\alpha| \leq 1$, $\alpha \in \mathbb{N}^n_{df} \setminus \{0\}$ in the dual (3.5) and allows its simplification (3.13) thanks to a property of the moment matrix described in [11]. Observe that the canonical form (3.10) of $\tilde{f}_d$ is *sparse* since $\tilde{f}_d$ differs from f in at most $3n$ entries only (recall that f has $\binom{n+df}{n}$ entries). This is another example of sparsity properties of optimal solutions of ℓ_1-norm minimization problems, already observed in other contexts (e.g., in some compressed sensing applications). Moreover, it has the following consequence for nonlinear 0/1 programs.

Corollary 3.5. *Let* $\mathbf{K} = \{0,1\}^n$ *and* $f \in \mathbb{R}[\mathbf{x}]$. *Then an optimal solution* $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]_{df}$ *of the inverse problem (3.1) for the* ℓ_1*-norm, is of the form*

$$\tilde{f}_d(\mathbf{x}) \,=\, f(\mathbf{x}) + \mathbf{b}'\mathbf{x}, \tag{3.14}$$

for some coefficient vector $\mathbf{b} \in \mathbb{R}^n$.

Proof. We briefly sketch the proof which is very similar to that of Theorem 3.4 even though $\mathbf{K}$ does not have a nonempty interior and df is not required to be even. For every $\alpha \in \mathbb{N}$, let $\bar{\alpha} \in \{0,1\}^n$ be such that $\bar{\alpha}_i = 1$ if $\alpha_i \neq 0$ and $\bar{\alpha}_i = 0$ otherwise. Then because of the boolean constraints $x_i^2 = x_i$, $i = 1, \ldots,$ (3.12) reads

$$\left\{\begin{array}{ll} \max\limits_{\mathbf{z}} & L_{\mathbf{z}}(f(0) - f)) \\ \text{s.t.} & \mathbf{M}_d(\mathbf{z}) \succeq 0; \quad L_{\mathbf{z}}(\mathbf{x}^\alpha) = L_{\mathbf{z}}(\mathbf{x}^{\bar{\alpha}}), \quad \alpha \in \mathbb{N}^n_{2d} \\ & |z_\alpha| \leq 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \{0\}. \end{array}\right.$$

But this combined with $\mathbf{M}_d(\mathbf{z}) \succeq 0$ implies that $\mathbf{M}_d(\mathbf{y})$ can be simplified to $\overline{\mathbf{M}}_d(\mathbf{z})$ with rows and columns indexed by only the square free monomials $\mathbf{x}^{\bar{\alpha}}$, $\alpha \in \mathbb{N}^n_{2d}$. Indeed, every non such column α is exactly identical to the column indexed by $\bar{\alpha}$. Also, it is relatively easy to show that $|L_{\mathbf{z}}(\mathbf{x}^{\bar{\alpha}})| \leq \max_i [L_{\mathbf{z}}(x_i)]$ for all $\bar{\alpha}$, and so (3.15) is equivalent to:

$$\left\{\begin{array}{ll} \max\limits_{\mathbf{z}} & L_{\mathbf{z}}(f(0) - f)) \\ \text{s.t.} & \overline{\mathbf{M}}_d(\mathbf{z}) \succeq 0; \quad L_{\mathbf{z}}(\mathbf{x}^\alpha) = L_{\mathbf{z}}(\mathbf{x}^{\bar{\alpha}}), \quad \alpha \in \mathbb{N}^n_{2d} \\ & \pm L_{\mathbf{z}}(\mathbf{x}_i) \,\leq\, 1, \qquad i = 1, \ldots, n. \end{array}\right. \tag{3.15}$$

Finally, let μ be a Borel measure with support exactly $\{0,1\}^n$, and scaled to satisfy $|\int x_i d\mu| < 1$, $i = 1, \ldots, n$. Its associated vector of moment $\mathbf{z} = (\int \mathbf{x}^\alpha d\mu)$, $\alpha \in \mathbb{N}^n$, is feasible in (3.15) and $\overline{\mathbf{M}}_d(\mathbf{z}) \succ 0$. Hence Slater's condition holds for (3.15), which in turn implies that there is no duality gap with its dual which reads:

$$\left\{\begin{array}{lll} \min\limits_{\mathbf{b}^1, \mathbf{b}^2, \lambda, \gamma} & \displaystyle\sum_{i=1}^n b_i^1 + b_i^2 & \\ \text{s.t.} & f - f(0) + (\mathbf{b}^1 - \mathbf{b}^2)'\mathbf{x} = \sigma_0 + \displaystyle\sum_{i=1}^n \sigma_i (x_i^2 - x_i) & \\ & \mathbf{b}^1, \mathbf{b}^2 \geq 0; \quad \sigma_0 \in \Sigma[\mathbf{x}]_d;\, \sigma_i \in \mathbb{R}[\mathbf{x}]_{d-1},\, j = 1, \ldots, n. & \end{array}\right.$$

Moreover, the dual has an optimal solution because the optimal value is bounded below by zero, and so $\tilde{f}_d$ is indeed of the form (3.14). □

3.4. **Structural constraints.** It may happen that the initial criterion $f \in \mathbb{R}[\mathbf{x}]$ has some structure that one wishes to keep in the inverse problem. For instance, in MAXCUT problems on $\mathbf{K} = \{-1,1\}^n$, f is a quadratic form $\mathbf{x} \mapsto \mathbf{x}'\mathbf{A}\mathbf{x}$ for some real symmetric matrix $\mathbf{A}$ associated with a graph (V, E), where $\mathbf{A}_{ij} \neq 0$ if and only if $(i,j) \in E$. Therefore, in the inverse optimization problem, one may wish that $\tilde{f}$ in (3.1) is also a quadratic form associated with the same graph (V, E), so that $\tilde{f}(\mathbf{x}) = \mathbf{x}'\tilde{\mathbf{A}}\mathbf{x}$ with $\tilde{\mathbf{A}}_{ij} = 0$ for all $(i,j) \notin E$.

So if $\Delta_f \subset \mathbb{N}^n_{df}$ denotes the subset of (structural) multi-indices for which f and $\tilde{f}$ should have same coefficient, then in (3.1) one includes the additional constraint $\tilde{f}_\alpha = f_\alpha$ for all $\alpha \in \Delta_f$. Notice that $0 \in \Delta_f$ because $\tilde{f}_0 = f_0$; see Remark 3.1. For instance, if $k = 1$ then (3.2) reads

$$
\rho^1_d := \min_{\tilde{f},\lambda_\alpha,\mathbf{Z}_j} \sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} \lambda_\alpha
$$

$$
\text{(3.16)} \qquad \begin{array}{ll}
\text{s.t.} & \lambda_\alpha + \tilde{f}_\alpha \geq f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \Delta_f \\
& \lambda_\alpha - \tilde{f}_\alpha \geq -f_\alpha, \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \Delta_f \\
& \langle \mathbf{Z}_0, \mathbf{B}_\alpha\rangle + \displaystyle\sum_{j=1}^m \langle \mathbf{Z}_j, \mathbf{C}^j_\alpha\rangle = \left\{\begin{array}{l} f_\alpha, \text{ if } 0 < \alpha \in \Delta_f \\ \tilde{f}_\alpha, \text{ if } \alpha \in \mathbb{N}^n_{df} \setminus \Delta_f \\ 0, \text{ if } \alpha = 0 \text{ or } |\alpha| > df \end{array}\right. \\
& \mathbf{Z}_j \succeq 0, \quad j = 0, 1, \ldots, m,
\end{array}
$$

and its dual has the equivalent form,

$$
\text{(3.17)} \qquad \left\{\begin{array}{ll}
\max\limits_{\mathbf{z}} & L_{\mathbf{z}}(f(0) - f)) \\
\text{s.t.} & \mathbf{M}_d(\mathbf{z}), \mathbf{M}_{d-v_j}(g_j\, \mathbf{z}) \succeq 0, \quad j = 1, \ldots, m, \\
& |z_\alpha| \leq 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \Delta_f.
\end{array}\right.
$$

In problems where df is even and Δ_f does not contain the monomials $\alpha \in \mathbb{N}^n_{df}$ such that $\mathbf{x}^\alpha = x_i^2$, or $\mathbf{x}^\alpha = x_i^{df}$, $i = 1, \ldots, n$, then $\tilde{f}$ has still the sepecial form described in Theorem 3.4, but with $b_k = 0$ if $\alpha = e_k \in \Delta_f$ (where all entries of e_k vanish except the one at position k).

3.5. **Asymptotics when $d \to \infty$.** We now relate $\mathbf{P}_d$, $d \in \mathbb{N}$, with the ideal inverse problem $\mathcal{P}$ in (2.8) when d increases.

Proposition 3.6. *Let $\mathbf{K}$ be as in (1.1) with nonempty interior. For every $k = 1, 2, \infty$, let $\tilde{f}_d \in \mathbb{R}[\mathbf{x}]_{df}$ (resp. $\tilde{f}^* \in \mathbb{R}[\mathbf{x}]_{df}$) be an optimal solution of (3.1) (resp. (2.8)) with associated optimal value ρ^k (resp. ρ^k).*

The sequence (ρ^k_d), $d \in \mathbb{N}$, is monotone nonincreasing and converges to $\hat{\rho}^k \geq \rho^k$. Moreover, every accumulation point $\hat{f} \in \mathbb{R}[\mathbf{x}]_{df}$ of the sequence $(\tilde{f}_d)$, $d \in \mathbb{N}$, is such that $\hat{f} - \hat{f}(0) \in \mathrm{Psd}_{df}(\mathbf{K})$ and $\|\hat{f} - f\|_k = \hat{\rho}^k$. Finally, if $\tilde{f}^* - \tilde{f}^*(0)$ is in $Q(g)$, then $\rho^k_d = \hat{\rho}^k = \rho^k$ for some d.

Proof. Observe that the sequence $(\tilde{f}_d)$, $d \in \mathbb{N}$, is contained in the ball $\{h : \|h - f\|_k \leq \rho^k_{d_0}\} \subset \mathbb{R}[\mathbf{x}]_{df}$, for some $d_0 \in \mathbb{N}$. So let $\hat{f}$ be an accumulation point of $(\tilde{f}_d)$. Since $\tilde{f}_d - \tilde{f}(0) \geq 0$ on $\mathbf{K}$ for all d, a simple continuity argument yields $\hat{f} - \hat{f}(0) \geq 0$ on $\mathbf{K}$, i.e., $\hat{f} - \hat{f}(0) \in \mathrm{Psd}_{df}(\mathbf{K})$. Moreover, the sequence (ρ^k_d) is obviously monotone nonincreasing and bounded below by zero. Hence $\lim_{d\to\infty} \rho^k_d =: \hat{\rho}^k \geq \rho^k$, and by continuity $\|\hat{f} - f\|_k = \hat{\rho}^k$.

Finally, if $\tilde{f}^* - \tilde{f}(0) \in Q(g)$ then $\tilde{f}^* - \tilde{f}(0) \in Q^{df}_d(g)$ for some d, and so $\tilde{f}^*$ is a feasible solution of (3.1) but with value $\rho^k \leq \rho^k_d$. Therefore, we conclude that $\tilde{f}^*$ is an optimal solution of (3.1). □

Proposition 3.6 relates ρ^k_d and ρ^k in a strong sense when $\tilde{f}^* - \tilde{f}(0) \in Q(g)$. However, we do know how restrictive is the assumption $\tilde{f}^* - \tilde{f}(0) \in Q(g)$ compared to $\tilde{f}^* - \tilde{f}(0) \in \mathrm{Psd}_{df}(\mathbf{K})$. Indeed, even though $\mathrm{Psd}_{df}(\mathbf{K}) = \mathrm{cl}\,(\cup_{\ell=0}^\infty) Q^{df}_\ell$ when $\mathbf{K}$ satisfies the assumptions of Theorem 2.7, in general an approximating sequence

$(f_\ell) \subset Q(g)$, $\ell \in \mathbb{N}$ (with $\|f_\ell - \tilde{f}^*\|_k \to 0$), does not have the property that $f_\ell(\mathbf{x}) - f_\ell(0) \geq 0$ for all $\mathbf{x}$ on $\mathbf{K}$. However, more can be said.

The dual of the ideal inverse problem $\mathcal{P}$. We now provide an explicit interpretation of the dual problems $\mathbf{P}_d^*$ in (3.5)-(3.6). Let $M(\mathbf{K})$ be the space of finite Borel measures on $\mathbf{K}$. Then obviously (3.5) is a relaxation of the following problem:

$$(3.18)\qquad \left\{ \begin{array}{ll} \mathbf{r}^1 = \displaystyle\max_{\mu\in M(\mathbf{K})} & \displaystyle\int_{\mathbf{K}} (f(0) - f(\mathbf{x}))\, d\mu(\mathbf{x}) \\ \text{s.t.} & \displaystyle\pm \int_{\mathbf{K}} \mathbf{x}^\alpha d\mu(\mathbf{x}) \,\leq\, 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \{0\}, \end{array} \right.$$

which, denoting by δ_0 the Dirac measure at $\mathbf{x} = 0$, and by $P(\mathbf{K})$ the space of Borel probability measures on $\mathbf{K}$, can be rewritten as

$$\left\{ \begin{array}{rl} \displaystyle\max_{\nu\in P(\mathbf{K}),\gamma\geq 0} & \gamma\,(\delta_0(f) - \nu(f)) \\ \text{s.t.} & \pm\gamma\,(\nu(\mathbf{x}^\alpha) - \delta_0(\mathbf{x}^\alpha)) \,\leq\, 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df}; \quad \nu(\mathbf{K}) = \gamma. \end{array} \right.$$

Similarly, (3.6) is a relaxation of the following problem:

$$(3.19)\qquad \mathbf{r}^\infty = \max_{\mu\in M(\mathbf{K})} \left\{ \int_{\mathbf{K}} (f(0) - f(\mathbf{x}))\, d\mu(\mathbf{x}) \;:\; \sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} \left| \int_{\mathbf{K}} \mathbf{x}^\alpha d\mu \right| \,\leq\, 1 \right\},$$

or, again, equivalently,

$$\max_{\nu\in P(\mathbf{K}),\gamma\geq 0} \left\{ \gamma\,(\delta_0(f) - \nu(f)) \;:\; \gamma \sum_{\alpha\in\mathbb{N}^n_{df}} |\nu(\mathbf{x}^\alpha) - \delta_0(\mathbf{x}^\alpha)| \,\leq\, 1; \quad \nu(\mathbf{K}) = \gamma \right\}.$$

Hence, in the dual problems (3.18) and (3.19) one searches for a finite Borel measure μ which concentrates as much as possible on the set $\{\mathbf{x} \in \mathbf{K} : f(\mathbf{x}) \leq f(0)\}$, and such that its moments up to order df are not too far from those of a measure supported at $\{0\} \in \mathbf{K}$.

In fact, and as one might have expected, (3.18) (resp. (3.19)) is the dual of $\mathcal{P}$ in (2.8) with $k = 1$ (resp. with $k = \infty$). For instance, with $k = 1$, to see that weak duality holds, let $\tilde{f} \in \mathbb{R}[\mathbf{x}]_{df}$ and $\mu \in M(\mathbf{K})$ be an arbitrary feasible solution of (2.8) and (3.18), respectively. Then:

$$\begin{aligned} \int_{\mathbf{K}} (f(0) - f)\, d\mu \quad &= \quad \underbrace{\int_{\mathbf{K}} (\tilde{f}(0) - \tilde{f}) d\mu}_{\leq 0} + \int_{\mathbf{K}} (\tilde{f} - f)\, d\mu \\ &\leq \quad \sum_{\alpha\in\mathbb{N}^n_{df}\setminus\{0\}} |\tilde{f}_\alpha - f_\alpha| \cdot \left| \int_{\mathbf{K}} \mathbf{x}^\alpha\, d\mu(\mathbf{x}) \right| \leq \|\tilde{f} - f\|_1, \end{aligned}$$

i.e., weak duality holds. We even have the following:

Lemma 3.7. *Let $\mathbf{K}$ in (1.1) be with nonempty interior and assume that $Q(g)$ is Archimedean. Let ρ^k be as in (2.8) with $k = 1, \infty$, and let $\mathbf{z}^d = (z^d_\alpha) \in \mathbb{R}^{s(d)}$ be a nearly optimal solution of (3.17), e.g., with value $L_{\mathbf{z}}(f(0) - f) \geq \rho^1_d - 1/d$, for all $d \in \mathbb{N}$.*

If $\liminf_{d\to\infty} z^d_0 < \infty$ then $\lim_{d\to\infty} \rho^k_d = \rho^k$ and (3.18) has an optimal solution $\mu^* \in M(\mathbf{K})$ which is supported on the set of global minimizers on $\mathbf{K}$ of the optimal solution $\tilde{f}^* \in \mathbb{R}[\mathbf{x}]_{df}$ of (2.8) (which contains $\{0\}$). Hence either $\rho^k = 0$ in which *case* 0 *is an optimal solution of* $\mathbf{P}$*, or* $\rho^k > 0$ *and* $\tilde{f}^*$ *has a another global minimizer* $\tilde{\mathbf{x}} \neq 0$ *on* $\mathbf{K}$ *with* $f(\tilde{\mathbf{x}}) < f(0)$*.*

Proof. The proof for the case $k = \infty$ is omitted as very similar to that of the case $k = 1$. Consider the subsequence d_i, $i \in \mathbb{N}$, such that $\liminf_{d\to\infty} z_0^d = \lim_{i\to\infty} z_0^{d_i} < \infty$. Using the Archimedean property (2.6) of $Q(g)$, we proceed exactly as in the proof of Theorem 3.2 in [9, p. 57–59]. There is a infinite sequence $\mathbf{z}^* = (z^*_\alpha)$, $\alpha \in \mathbb{N}^n$, and a subsequence (still denoted d_i for notational convenience), such that for every $\alpha \in \mathbb{N}^n$, $z_\alpha^{d_i} \to z^*_\alpha$. Moreover, from the convergence $\mathbf{z}^{d_i} \to \mathbf{z}^*$, $\mathbf{M}_d(g_j\,\mathbf{z}^*) \succeq 0$ for every $d \in \mathbb{N}$ and every $j = 0, 1, \ldots, m$; hence by Putinar's Theorem [12], $\mathbf{z}^*$ is the sequence of moments of a finite Borel measure μ^* supported on $\mathbf{K}$. Moreover, since $|z_\alpha^{d_i}| \leq 1$ for all $\alpha \in \mathbb{N}^n_{df} \setminus \{0\}$, we obtain

$$|\int_{\mathbf{K}} \mathbf{x}^\alpha\, d\mu^*| \,=\, |z^*_\alpha| = \lim_{i\to\infty} |z_\alpha^{d_i}| \,\leq\, 1 \quad \forall \alpha \in \mathbb{N}^n_{df} \setminus \{0\},$$

which proves that μ^* is feasible for (3.18). Finally, by monotonicity of the sequence (ρ^1_d), $d \in \mathbb{N}$, and using $\rho^1_d \geq \mathbf{r}^1$ for all d,

$$\begin{aligned} \mathbf{r}^1 \leq \lim_{d\to\infty} \rho^1_d &= \lim_{i\to\infty} \rho^1_{d_i} \,=\, \lim_{i\to\infty} L_{\mathbf{z}^{d_i}}(f(0) - f) \\ &= L_{\mathbf{z}^*}(f(0) - f) = \int_{\mathbf{K}} (f(0) - f) d\mu^*, \end{aligned}$$

which proves that μ^* is an optimal solution of (3.18). Observe that $\mu^*(\mathbf{K}) < M$ because $\mu^*(\mathbf{K}) = z^*_0 = \lim_{i\to\infty} z_0^{d_i} < M$.

We next prove that $\rho^1 = \mathbf{r}^1$. Since $\liminf_{d\to\infty} z_0^d < M$, consider the linear program :

$$\left\{ \begin{array}{ll} \displaystyle\max_{\mu\in M(\mathbf{K})} & \displaystyle\int_{\mathbf{K}} (f(0) - f(\mathbf{x}))\, d\mu(\mathbf{x}) \\ \text{s.t.} & \displaystyle\pm\int_{\mathbf{K}} \mathbf{x}^\alpha d\mu(\mathbf{x}) \,\leq\, 1, \quad \forall\, \alpha \in \mathbb{N}^n_{df} \setminus \{0\};\ \mu(\mathbf{K}) \leq M, \end{array} \right. \tag{3.20}$$

which is (3.18) with the additional constraint $\mu(\mathbf{K}) \leq M$. From what precedes, its optimal value is still $\mathbf{r}^1$, and μ^* is an optimal solution. Its dual reads

$$\theta = \min_{\gamma\geq 0, \tilde{f}\in\mathbb{R}[\mathbf{x}]_{df}} \{M\gamma + \|\tilde{f} - f\|_1 \,:\, \tilde{f} - \tilde{f}(0) + \gamma \in \mathrm{Psd}(\mathbf{K})\}, \tag{3.21}$$

Observe that to any feasible solution $\tilde{f}$ of (2.8) corresponds the feasible solution $(0, \tilde{f})$ of (3.21) with same value $\|\tilde{f} - f\|_1$. Hence $\theta \leq \rho^1$. But there is no duality gap between (3.20) and (3.21) because $\mathbf{K}$ is compact and the constraint $\mu(\mathbf{K}) \leq M$ makes the feasible set bounded. Therefore,

$$\rho^1 \,\geq\, \theta \,=\, \mathbf{r}^1 \,=\, \int_{\mathbf{K}} (f(0) - f)\, d\mu^* \,\leq\, \rho^1,$$

i.e., $\rho^1 = \mathbf{r}^1$. Finally, since $\tilde{f}^*(0) = f(0)$,

$$\begin{aligned} \rho^1 = \int_{\mathbf{K}} (f(0) - f)\, d\mu^* &= \underbrace{\int_{\mathbf{K}} (\tilde{f}^*(0) - \tilde{f}^*)\, d\mu^*}_{\leq 0} + \int_{\mathbf{K}} (\tilde{f}^* - f)\, d\mu^* \\ &\leq \|\tilde{f}^* - f\|_1 = \rho^1, \end{aligned}$$

which implies $\mu^*(\{\mathbf{x}\,:\,\tilde{f}^*(\mathbf{x})-\tilde{f}^*(0)>0\})=0$, that is, the support of μ^* is contained in the set of global minimizers of $\tilde{f}^*$ (which contains $\{0\}$). Therefore, if $\rho^1>0$ then necessarily there is another global minimizer $0\neq\tilde{\mathbf{x}}\in\mathbf{K}$ of $\tilde{f}^*$ with $f(\tilde{\mathbf{x}})<f(0)$, otherwise $\rho^1=\int(f(0)-f)d\mu^*=0$.

□

3.6. **Convexity.** One may wish to restrict to search for convex polynomials $\tilde{f}\in\mathbb{R}[\mathbf{x}]_{df}$ (no matter if f itself is convex). For instance if the g_j's are concave (so that $\mathbf{K}$ is convex) but f is not, one may wish to find the convex optimization problem whose $\mathbf{y}\in\mathbf{K}$ is an optimal solution and with convex polynomial criterion $\tilde{f}\in\mathbb{R}[\mathbf{x}]_{df}$ closest to f.

If $df>2$ then in (3.1) it suffices to add the additional Putinar's certificate

$$(\mathbf{x},\mathbf{u})\mapsto\;\mathbf{u}^T\nabla^2\tilde{f}(\mathbf{x})\,\mathbf{u}\;=\;\sum_{j=0}^m\psi_j(\mathbf{x},\mathbf{u})\,g_j(\mathbf{x})+\psi_{m+1}(\mathbf{x},\mathbf{u})(1-\|\mathbf{u}\|^2),\tag{3.22}$$

with $\psi_{m+1}\in\mathbb{R}[\mathbf{x},\mathbf{u}]$ and $\psi_j\in\Sigma_{d-v_j}[\mathbf{x},\mathbf{u}]$, for all $j=0,1,\ldots,m$. Indeed, (3.22) is a Putinar's certificate of convexity for $\tilde{f}$ on $\mathbf{K}$, with degree bound d. As the coefficients of the polynomial $(\mathbf{x},\mathbf{u})\mapsto\;\mathbf{u}^T\nabla^2\tilde{f}(\mathbf{x})\mathbf{u}$ are linear in the coefficients of $\tilde{f}$, (3.22) will translate into additional semidefinite constraints in (3.2).

If $df\leq 2$, i.e. if $f(\mathbf{x})=\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x}+c$ for some real symmetric matrix $\mathbf{A}\in\mathbb{R}^{n\times n}$, some vector $\mathbf{b}\in\mathbb{R}^n$ and some scalar $c\in\mathbb{R}$, then $\tilde{f}(\mathbf{x})=\frac{1}{2}\mathbf{x}^T\tilde{\mathbf{A}}\mathbf{x}+\tilde{\mathbf{b}}^T\mathbf{x}+\tilde{c}$ for some real symmetric matrix $\tilde{\mathbf{A}}\in\mathbb{R}^{n\times n}$, some $\tilde{\mathbf{b}}\in\mathbb{R}^n$ and some $\tilde{c}\in\mathbb{R}$. In that case, in (3.1) it suffices to add constraint $\nabla^2\tilde{f}(\mathbf{x})=\tilde{\mathbf{A}}\succeq 0$, which is just a Linear Matrix Inequality (LMI). And therefore, again, (3.1) can be rewritten as a semidefinite program, namely (3.2)-(3.4) with the additional LMI constraint $\tilde{\mathbf{A}}\succeq 0$.

Notice that for $k=1,2$, it also makes sense to search for $\tilde{f}\in\mathbb{R}[\mathbf{x}]_2$ even if f has degree $df>2$, i.e., if $f(\mathbf{x})=c+\mathbf{b}^T\mathbf{x}+\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+h(\mathbf{x})$ where $h\in\mathbb{R}[\mathbf{x}]$ does not contains monomials of degree smaller than 3. This means that one searches for the convex program with quadratic cost closest to f.

So for instance, in the case where one searches for $\tilde{f}\in\mathbb{R}[\mathbf{x}]_2$, and given $\mathbf{y}\in\mathbf{K}$ let $J(\mathbf{y}):=\{j\in\{1,\ldots,m\}\,:\,g_j(\mathbf{y})=0\}$ be the set of constraints that are active at $\mathbf{y}$. If the g_j's that define $\mathbf{K}$ are concave then one may simplify (3.1). Writing $\tilde{f}=\frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}+\mathbf{b}^T\mathbf{x}+c$, and with $k=1,2$, (3.1) now reads:

$$\begin{aligned}\rho^k:=\min_{\tilde{\mathbf{A}},\tilde{\mathbf{b}},\lambda}\quad&\|f-\tilde{f}\|_k\\ \text{s.t.}\quad&\tilde{\mathbf{A}}\,\mathbf{y}+\tilde{\mathbf{b}}\;=\;\sum_{j\in J(\mathbf{y})}\lambda_j\,\nabla g_j(\mathbf{y})\\ &\tilde{\mathbf{A}}\succeq 0;\;\lambda_j\geq 0,\quad j\in J(\mathbf{y}).\end{aligned}$$

So, as we did in the previous section, and possibly after the change of variable $\mathbf{x}':=\mathbf{x}-\mathbf{y}$, with no loss of generality one may and will assume that $\mathbf{y}=0$, in which case (3.23) simplifies to

$$\begin{aligned}\rho^k:=\min_{\tilde{\mathbf{A}},\tilde{\mathbf{b}},\lambda}\quad&\|f-\tilde{f}\|_k\\ \text{s.t.}\quad&\tilde{\mathbf{b}}\;=\;\sum_{j\in J(0)}\lambda_j\,\nabla g_j(0)\\ &\tilde{\mathbf{A}}\succeq 0;\;\lambda_j\geq 0,\quad j\in J(0),\end{aligned}\tag{3.23}$$

which in turn simplifies to

$$(3.24)\qquad \begin{array}{rcl} \rho^1 &=& \min_{\tilde{\mathbf{A}}\succeq 0} \|\tilde{\mathbf{A}}-\mathbf{A}\|_1 + \min_{\lambda\geq 0} \|\mathbf{b} - \sum_{j\in J(0)} \lambda_j \nabla g_j(0)\|_1 \\ \rho^\infty &=& \sup \left[\min_{\tilde{\mathbf{A}}\succeq 0} \|\tilde{\mathbf{A}}-\mathbf{A}\|_\infty \,,\, \min_{\lambda\geq 0} \|\mathbf{b} - \sum_{j\in J(0)} \lambda_j \nabla g_j(0)\|_\infty \right]. \end{array}$$

Observe that (3.24) can be solved in two steps. One first solves the problem $\min_{\lambda\geq 0} \|\mathbf{b}-\sum_{j\in J(0)} \lambda_j \nabla g_j(0)\|_k$, which is a linear program with finite value, hence with an optimal solution. One next solves the problem $\min_{\tilde{\mathbf{A}}\succeq 0} \|\tilde{\mathbf{A}}-\mathbf{A}\|_k$ which computes the ℓ_k-projection of $\mathbf{A}$ onto the closed convex cone of positive semidefinite matrices (a semidefinite program with an optimal solution).

Lemma 3.8. *Let $\mathbf{K}\subset\mathbb{R}^n$ be as in (1.1) with g_j being concave for every $j=1,\ldots,m$. Then (3.23) has an optimal solution $\tilde{f}^*\in\mathbb{R}[\mathbf{x}]_2$ and 0 is an optimal solution of the convex optimization problem $\mathbf{P}'$: $\min\{\tilde{f}^*(\mathbf{x}) : \mathbf{x}\in\mathbf{K}\}$.*

Proof. Let $(\tilde{f},\lambda)$ (with $\tilde{f}\in\mathbb{R}[\mathbf{x}]_2$) be any feasible solution of (3.23). The constraint in (3.23) states that $\nabla L(0)=0$, where $L\in\mathbb{R}[\mathbf{x}]$ is the Lagrangian polynomial $\mathbf{x}\mapsto L(\mathbf{x}) := \tilde{f}(\mathbf{x}) - \sum_{j\in J(0)} \lambda_j\, g_j(\mathbf{x})$, which is convex on $\mathbf{K}$ because the g_j's are concave, the λ_j's are nonnegative, and $\tilde{f}$ is convex. Therefore $\nabla L(0)=0$ implies that 0 is a global minimum of L on $\mathbb{R}^n$ and a global minimum of $\tilde{f}$ on $\mathbf{K}$ because

$$(3.25)\qquad \tilde{f}(\mathbf{x}) \,\geq\, L(\mathbf{x}) \,\geq\, L(0) \,=\, \tilde{f}(0), \qquad \forall\, \mathbf{x}\in\mathbf{K}.$$

It remains to prove that (3.23) has an optimal solution $\tilde{f}^*$. But we have seen that (3.23) is equivalent to (3.24) for which an optimal solution can be found by solving a linear program and a semidefinite program. □

So in this case where the g_j's are concave (hence $\mathbf{K}$ is convex), one obtains the convex programming problem with quadratic cost, whose criterion is the closest to f for the ℓ_k-norm.

3.7. **First-order inverse local optimality condition.** Finally we also consider a test of first-order "inverse local optimality". Again, let $J(\mathbf{y})$ identify the constraints that are active at some current iterate $\mathbf{y}\in\mathbf{K}$, and with $k=1,2$ or $k=\infty$, consider the optimization problem:

$$(3.26)\qquad \begin{array}{rll} \rho_d := & \min_{\lambda_j\in\mathbb{R}^{|J(\mathbf{y})|},\tilde{f}\in\mathbb{R}[\mathbf{x}]_{df}} & \|f-\tilde{f}\|_k \\ & \text{s.t.} & \nabla\tilde{f}(\mathbf{y}) = \sum_{j\in J(\mathbf{y})}^{m} \lambda_j \,\nabla\, g_j(\mathbf{y}) \\ & & \lambda_j\geq 0, \quad j\in J(\mathbf{y}) \end{array}$$

which can be transformed into a linear program if $k=1$ or $k=\infty$. Indeed, the KKT-constraints

$$\nabla\tilde{f}(\mathbf{y}) = \sum_{j\in J(\mathbf{y})}^{m} \lambda_j \,\nabla\, g_j(\mathbf{y})$$

yield n linear constraints on the coefficients $(\tilde{f}_\alpha)\in\mathbb{R}^{s(df)}$ and $\lambda\in\mathbb{R}^{|J(\mathbf{y})|}$, and as in (3.2) or (3.3), one replaces $\min\|\tilde{f}-f\|_k$ with an appropriate linear criterion and appropriate linear constraints.

Obviously, if $\tilde{f}_d$ is an optimal solution of the linear program (3.26) then $\mathbf{y} \in \mathbf{K}$ satisfies the KKT-optimality conditions for the problem $\min_{\mathbf{x}}\{\tilde{f}_d(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$.

3.8. **Linear Programming.** Consider the linear program $\mathbf{P} = \min\{\mathbf{c}'\mathbf{x} \,:\, \mathbf{A}\mathbf{x} = \mathbf{b};\, \mathbf{x} \geq 0\}$ for some $\mathbf{A} \in \mathbb{R}^{m\times n}$ and $\mathbf{c} \in \mathbb{R}^n$. Suppose that $\mathbf{y} \in \mathbb{R}^n$ is vertex of the corresponding polyhedron and let $\mathbf{A}_\sigma$ be the square matrix associated with a basis $\sigma \subset \{1,\ldots,n\}$ of $|\sigma| = m$ columns of $\mathbf{A}$. Let $\lambda := (\mathbf{A}'_\sigma)^{-1}\mathbf{c}_\sigma$ where $\mathbf{c}_\sigma = (c_j)$, $j \in \sigma$, and let δ_j be the reduced costs associated with a column $j \notin \sigma$, that is, $\delta_j := c_j - \mathbf{c}'_\sigma \mathbf{A}_\sigma^{-1}\mathbf{A}_j$. Then :

Lemma 3.9. *Let $\mathbf{y} \in \mathbb{R}^n$ be a vertex of the polyhedron $\{\mathbf{x} \in \mathbb{R}^n_+ \,:\, \mathbf{A}\mathbf{x} = \mathbf{b}\}$, associated with a basis σ. Then $\mathbf{y}$ is an optimal solution of the new linear program* $\max\{\tilde{\mathbf{c}}'\mathbf{x} \,:\, \mathbf{A}\mathbf{x} = \mathbf{b};\, \mathbf{x} \geq 0\}$ *where*

$$\tilde{\mathbf{c}}_\sigma = \begin{cases} c_j & \text{if } j \in \sigma \text{ or if } j \notin \sigma \text{ and } \delta_j \leq 0, \\ c_j - \delta_j \,(= \lambda' \mathbf{A}_\sigma^{-1}\mathbf{A}_j) & \text{if } j \notin \sigma \text{ and } \delta_j > 0, \end{cases}$$

and $\rho_d = \sum_{\delta_j > 0} \delta_j$.

Proof. With the new cost vector $\tilde{\mathbf{c}}$ of Lemma 3.9 the optimality conditions are satisfied at the vertex $\mathbf{y}$, any other such cost $\hat{\mathbf{c}}$ would produce a larger ℓ_1-norm $\|\mathbf{c} - \hat{\mathbf{c}}\|_1$. □

3.9. **Illustrative examples and discussion.** We here provide some simple illustrative examples and show that the representation of the set $\mathbf{K}$ may be important for getting a Putinar certificate faster.

Example 1. Let $n = 2$ and consider the optimization problem $\mathbf{P} : f^* = \min_{\mathbf{x}}\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ with $\mathbf{x} \mapsto f(\mathbf{x}) = x_1 + x_2$, and

$$\mathbf{K} = \{\mathbf{x} \in \mathbb{R}^2 \,:\, x_1x_2 \geq 1;\, 1/2 \leq \mathbf{x} \leq 2\,\}.$$

The polynomial f is convex and the set $\mathbf{K}$ is convex as well, but the polynomials that define $\mathbf{K}$ are not all concave. That is, $\mathbf{P}$ is a convex optimization problem, but not a convex programming problem. The point $\mathbf{y} = (1,1) \in \mathbf{K}$ is a global minimizer and the KKT conditions at $\mathbf{y}$ are satisfied, i.e., $\lambda = (1,0,0,0,0) \in \mathbb{R}^4$,

$$\nabla f(\mathbf{x}) - \lambda_1 \nabla g_1(\mathbf{x}) = 0 \text{ with } \mathbf{x} = (1,1) \text{ and } \lambda_1 = 1.$$

However, the Lagrangian

$$\mathbf{x} \mapsto L(\mathbf{x}) := f(\mathbf{x}) - f^* - \lambda_1\, g_1(\mathbf{x}) = x_1 + x_2 - 1 - x_1x_2,$$

is not convex and so $(1,1)$ is not a global minimum of L on $\mathbb{R}^2$. This example just illustrates the fact that even in the convex case where the g_j's are not concave, the KKT conditions do not provide a *certificate* of global optimality, contrary to "convex programming" where since L is now convex, obviously using $L(\mathbf{x}) \geq L(\mathbf{y}) = 0$ (because $\nabla L(\mathbf{y}) = 0$),

$$f(\mathbf{x}) - f^* \geq L(\mathbf{x}) \geq L(\mathbf{y}) = 0,$$

whenever $\mathbf{x} \in \mathbf{K}$, and so $f(\mathbf{x}) \geq f^*$ for all $\mathbf{x} \in \mathbf{K}$, the desired certificate of global optimality.

Next, if we now use the test of inverse optimality with $d = 1$, one searches for a polynomial $\tilde{f}_d$ of degree at most $df = 1$, and such that

$$\tilde{f}_d(\mathbf{x}) - \tilde{f}_d(1,1) = \sigma_0(\mathbf{x}) + \sigma_1(\mathbf{x})(x_1x_2 - 1) + \sum_{i=1}^{2} \psi_i(\mathbf{x})(2 - x_i) + \phi_i(\mathbf{x})(x_i - 1/2),$$

for some s.o.s. polynomials $\sigma_1, \psi_i\, \phi_i \in \Sigma[\mathbf{x}]_0$ and some s.o.s. polynomial $\sigma_0 \in \Sigma[\mathbf{x}]_1$. But then necessarily $\sigma_1 = 0$ and ψ_i, ϕ_i are constant, which in turn implies that σ_0 is a constant polynomial. A straightforward calculation shows that $\tilde{f}_1(\mathbf{x}) = 0$ for all $\mathbf{x}$, and so $\rho_1^1 = 2$. And indeed this is confirmed when solving[3] (3.5) with $d = 1$. Solving again (3.5) with now $d = 2$ yields $\rho_2^1 = 2$ (no improvement) and with $d = 3$ we obtain the desired result $\rho_3^1 = 0$.

On the other hand, if now $\mathbf{K}$ has the representation:

$$\{\mathbf{x} \,:\, x_1x_2 - 1 \,\geq\, 0; \quad (x_i - 1/2)(2 - x_i) \geq\, 0; \quad i = 1, 2\,\},$$

then the situation differs because in fact

$$x_1 + x_2 - 2 \,=\, \frac{1}{5} + \frac{2}{5}(x_1 - x_2)^2 + \frac{4}{5}(x_1x_2 - 1) + \frac{2}{5}\sum_{i=1}^{2}(x_i - 1/2)(2 - x_i),$$

i.e., $f - f^*$ has a Putinar's certificate with degree bound $d = 1$. Hence the test of inverse optimality yields $\rho_1^1 = 0$ with $\tilde{f}_1 = f$.

The above example illustrates that the representation of $\mathbf{K}$ may be important.

Example 2. Again consider Example 1 but now with $\mathbf{y} = (1.1, 0.909) \in \mathbf{K}$, which is not a global optimum of f on $\mathbf{K}$ any more. By solving (3.5) with $d = 1$ we still find $\rho_1^1 = 0$, and with $d = 2$ we find $\tilde{f}_2(\mathbf{x}) \approx 0.82782x_1 + x_2$. And indeed by solving (using GloptiPoly) the new optimization problem with criterion $\tilde{f}_2$ we find the global minimizer $(1.0991, 0.9098) \approx \mathbf{y}$.

With $\mathbf{y} = (1.8, 0.5556) \in \mathbf{K}$, we find $\tilde{f}_2(\mathbf{x}) \approx 0.3097x_1 + x_2$. By solving the new optimization problem with criterion $\tilde{f}_2$ we find the global minimizer $(1.797, 0.5565) \approx \mathbf{y}$.

Example 3. Consider the MAXCUT problem $\max\{\mathbf{x}'\mathbf{A}\mathbf{x} \,:\, \mathbf{x}_i^2 = 1, i = 1, \ldots, n\}$ where $\mathbf{A} = \mathbf{A}' \in \mathbb{R}^{n\times n}$ and $\mathbf{A}_{ij} = 1/2$ for all $i \neq j$. For n odd, an optimal solution is $\mathbf{y} = (y_j)$ with $y_j = 1$, $j = 1, \ldots \lceil n/2 \rceil$, and $y_j = -1$ otherwise. However, the first semidefinite relaxation

$$\max\,\{\lambda \,:\, \mathbf{x}'\mathbf{A}\mathbf{x} - \lambda = \sigma + \sum_{j=1}^{n}\gamma_i(x_i^2 - 1);\; \sigma \in \Sigma[\mathbf{x}]_1;\; \lambda, \gamma \in \mathbb{R}\}$$

provides the lower bound $-n/2$ (with famous Goemans-Williamson ratio guarantee). So $\mathbf{y}$ cannot be obtained from the first semidefinite relaxation even though it is an optimal solution. The inverse optimization problem reads: Find the quadratic form $\mathbf{x} \mapsto \mathbf{x}'\tilde{\mathbf{A}}\mathbf{x}$ such that $\mathbf{x}'\tilde{\mathbf{A}}\mathbf{x} - \mathbf{y}'\tilde{\mathbf{A}}\mathbf{y} \,=\, \sigma + \sum_{j=1}^{n}\gamma_i(x_i^2 - 1)$, for some $\sigma \in \Sigma[\mathbf{x}]_1$, $\lambda, \gamma \in \mathbb{R}$, and which minimizes the ℓ_1-norm $\|\mathbf{A} - \tilde{\mathbf{A}}\|_1$. This is an inverse optimization problem with structural constraints as described in Section 3.4 (since we search for a quadratic form and not an arbitrary quadratic polynomial

[3]To solve (3.5) we have used the GloptiPoly software of Henrion et al. [4], and dedicated to solving the Generalized Problem of Moments whose problem (3.5) is only a special case.

$\tilde{f}_2$). Hence, solving (3.16) for $n = 5$ with $\mathbf{y}$ as above, we find that

$$\tilde{\mathbf{A}} = \frac{1}{2}\left[\begin{array}{ccccc} 0 & 2/3 & 2/3 & 1 & 1 \\ 2/3 & 0 & 2/3 & 1 & 1 \\ 2/3 & 2/3 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{array}\right],$$

that is, only the entries $(i, j) \in \{(1, 2), (1, 3), (2, 3)\}$ are modified from 1/2 to 1/3.

Conclusion

We have presented a paradigm for inverse polynomial optimization. Crucial is Putinar's Positivstellensatz which provides us with the desired certificate of global optimality for a given feasible point $\mathbf{y} \in \mathbf{K}$ and a candidate criterion $\tilde{f}$. In addition, to some extent, the size of the certificate can be adapted to the computational capabilities available. Finally, and remarkably, when using the ℓ_1-norm the resulting inverse optimal criterion $\tilde{f}$ has a simple and explicit *canonical* form. We hope that the concept of inverse optimization will receive more attention from the optimization community as it could even provide an alternative stopping criterion at the current iterate $\mathbf{y} \in \mathbf{K}$ of any local optimization algorithm for solving the original problem $\mathbf{P}$.

References

[1] S. Ahmed and Yongpei Guan. The inverse optimal value problem. Math. Program. Ser. A **102**, 91–110 (2005)

[2] R.K. Ahuja and J.B. Orlin. Inverse optimization, Oper. Res. **49**, 771–783 (2001)

[3] D. Burton and P. Toint. On an instance of the inverse shortest path problem, Math. Program. **63**, 45–61 (1992)

[4] D. Henrion, J.B. Lasserre, J. Lofberg. Gloptipoly 3: moments, optimization and semidefinite programming, Optim. Methods and Softw. **24**, 761–779 (2009)

[5] C. Heuberger. Inverse combinatorial oprtimization: A survey on problems, methods and results, Journal of Comb. Optim. **8**, 329–361 (2004).

[6] S. Huang, Z. Liu. On the inverse problem of linear programming and its application to minimum weight perfect k-matching, Eur. J. Oper. Res. **112**, pp. 421–426 (199)

[7] R.A. Freeman and P.V. Kokotovic. Inverse optimality in robust stabilization, SIAM J. Control Optim. **34**, pp. 1365–1391 (1996)

[8] J.B. Lasserre. Global optimization with polynomials and the problem of moments, SIAM J. Optim. **11**, pp. 796–817 (2001)

[9] J.B. Lasserre. Moments and sums of squares for polynomial optimization and related problems, J. Glob. Optim. **45**, 39–61 (2009)

[10] J.B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity, SIAM J. Optim. **17**, pp. 822–843 (2006)

[11] J.B. Lasserre, T. Netzer. SOS approximations of nonnegative polynomials via simpla high degree perturbation, Math. Z. **256**, pp. 99–112 (2006)

[12] M. Putinar. Positive polynomials on compact semi-algebraic sets, Indiana Univ. Math. J. **42**, pp. 969–984 (1993)

[13] A. Schaefer. Inverse integer programming, Optim. Letters. **3**, 483–489 (2004)

[14] H. Waki, S. Kim, M. Kojima, M. Maramatsu. Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity, SIAM J. Optim. **17**, pp. 218–242 (2006)

[15] H. Wolkowicz, R. Saigal, L. Vandenberghe. *Handbook of Semidefinite Programming: Theory, Algorithms, and Applications*, Kluwer Academic Publishers, Boston, MA (2000)

[16] L.A. Wolsey. Integer programming duality: Price functions and sensitivity analysis, Math. Prog. **20**, 173–195 (1981)

[17] J. Zhang, Z. Liu. Calculating some inverse linear programming problems, J. Comp. Appl. Math. **72**, pp. 261–273 (1996)
[18] J. Zhang, Z. Liu. A further study on inverse linear programming problems, J. Comp. Appl. Math. **106**, pp. 345–359 (1999)
[19] J. Zhang, Z. Ma, C. Yang. A column generation method for inverse shortest path problems, Math. Methods Oper. Res. **41**, pp. 347–358 (1995)

LAAS-CNRS and Institute of Mathematics, University of Toulouse, LAAS, 7 avenue du Colonel Roche, 31077 Toulouse Cédex 4,France
E-mail address: lasserre@laas.fr