A Lagrangian relaxation view of linear and semidefinite hierarchies

Jean-Bernard Lasserre

▶ **To cite this version:**

Jean-Bernard Lasserre. A Lagrangian relaxation view of linear and semidefinite hierarchies. SIAM Journal on Optimization, 2013, 23, pp.1742–1756. hal-00742918

HAL Id: hal-00742918
https://hal.science/hal-00742918

Submitted on 17 Oct 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

A Lagrangian relaxation view of linear and semidefinite hierarchies

Jean B. Lasserre*

Abstract

We consider the general polynomial optimization problem $\mathbf{P}: f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$ where $\mathbf{K}$ is a compact basic semi-algebraic set. We first show that the standard Lagrangian relaxation yields a lower bound as close as desired to the global optimum f^*, *provided* that it is applied to a problem $\tilde{\mathbf{P}}$ equivalent to $\mathbf{P}$, in which sufficiently many redundant constraints (products of the initial ones) are added to the initial description of $\mathbf{P}$. Next we show that the standard hierarchy of LP-relaxations of $\mathbf{P}$ (in the spirit of Sherali-Adams' RLT) can be interpreted as a brute force simplification of the above Lagrangian relaxation in which a nonnegative polynomial (with coefficients to be determined) is replaced with a constant polynomial equal to zero. Inspired by this interpretation, we provide a systematic improvement of the LP-hierarchy by doing a much less brutal simplification which results into a parametrized hierarchy of semidefinite programs (and not linear programs any more). For each semidefinite program in the parametrized hierarchy, the semidefinite constraint has a fixed size $O(n^k)$, independently of the rank in the hierarchy, in contrast with the standard hierarchy of semidefinite relaxations. The parameter k is to be decided by the user. When applied to a non trivial class of convex problems, the first relaxation of the parametrized hierarchy is exact, in contrast with the LP-hierarchy where convergence cannot be finite. When applied to 0/1 programs it is at least as good as the first one in the hierarchy of semidefinite relaxations. However obstructions to exactness still exist and are briefly analyzed. Finally, the standard semidefinite hierarchy can also be viewed as a simplification of an extended Lagrangian relaxation, but different in spirit as sums of squares (and not scalars) multipliers are allowed.

Keywords: Global and 0/1 optimization; approximation algorithms; linear and semidefinite relaxations; Lagrangian relaxations.

*LAAS-CNRS and Institute of Mathematics, University of Toulouse, 7 Avenue du Colonel Roche, 31077 Toulouse, France. lasserre@laas.fr

1 Introduction

Recent years have seen the development of (global) semi-algebraic optimization and in particular LP- or semidefinite relaxations for the polynomial optimization problem:

$$\mathbf{P}: \quad f^* = \min_{\mathbf{x}} \{ f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{K} \} \tag{1.1}$$

where $f \in \mathbb{R}[\mathbf{x}]$ is a polynomial and $\mathbf{K} \subset \mathbb{R}^n$ is the basic semi-algebraic set

$$\mathbf{K} \,=\, \{ \mathbf{x} \in \mathbb{R}^n \,:\, g_j(\mathbf{x}) \,\geq\, 0, \quad j = 1, \ldots, m \}, \tag{1.2}$$

for some polynomials $g_j \in \mathbb{R}[\mathbf{x}]$, $j = 1, \ldots, m$.

In particular, associated with $\mathbf{P}$ are two hierarchies of convex relaxations:

- *Semidefinite* relaxations based on Putinar's certificate of positivity on $\mathbf{K}$ [16], where the d-th convex relaxation of the hierarchy is a semidefinite program which solves the optimization problem

$$\gamma_d = \max_{t, \sigma_j} \{ \, t \,:\, f - t = \sigma_0 + \sum_{j=1}^{n} \sigma_j \, g_j \}. \tag{1.3}$$

The unknowns σ_j are sums of squares polynomials with the degree bound constraint $\deg \sigma_j g_j \leq 2d$, $j = 0, \ldots, m$, and the expression in (1.3) is a certificate of positivity on $\mathbf{K}$ for the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - t$.

- *LP*-relaxations based on Krivine-Stengle's certificate of positivity on $\mathbf{K}$ [9, 19], where the d-th convex relaxation of the hierarchy is a linear program which solves the optimization problem

$$\theta_d \,=\, \max_{\lambda \geq 0, t} \{ t \,:\, f - t \;\; = \;\; \sum_{(\alpha,\beta) \in \mathbb{N}^{2m}_d} \lambda_{\alpha\beta} \left(\prod_{j=1}^{m} g_j^{\alpha_j} \right) \times \left(\prod_{j=1}^{m} (1 - g_j)^{\beta_j} \right) \Bigg\}, \tag{1.4}$$

where $\mathbb{N}^{2m}_d = \{ (\alpha, \beta) \in \mathbb{N}^{2m} : \sum_j \alpha_j + \beta_j \leq d \}$. The unknown are t and the nonnegative scalars $\lambda = (\lambda_{\alpha\beta})$, and it is assumed that $0 \leq g_j \leq 1$ on $\mathbf{K}$ (possibly after scaling) and the family $\{ g_i, 1 - g_i \}$ generates the algebra

$\mathbb{R}[\mathbf{x}]$ of polynomials. Problem (1.4) is an LP because stating that the two polynomials in both sides of "=" are equal yields linear constraints on the $\lambda_{\alpha\beta}$'s. For instance, the LP-hierarchy from Sherali-Adams' RLT [17] and their variants [18] are of this form. See more details in §3.3.

In both cases, (γ_d) and (θ_d), $d \in \mathbb{N}$, provide two monotone nondecreasing sequences of lower bounds on f^* and if $\mathbf{K}$ is compact then both converge to f^* as one let d increase. For more details as well as a comparison of such relaxations the interested reader is referred to e.g. Lasserre [12, 10] and Laurent [13], as well as Chlamtac and Tulsiani [5] for the impact of LP- and SDP-hierarchies on approximation algorithms in combinatorial optimization.

Of course, in principle, one would much prefer to solve LP-relaxations rather than semidefinite relaxations (i.e. compute θ_d rather than γ_d) because present LP-software packages can solve problems with millions of variables and constraints, which is far from being the case for semidefinite solvers. And so the hierarchy (1.3) applies to problems of modest size only unless some sparsity or symmetry is taken into account in which case specialized variants can handle problems of much larger size. However, on the other hand, the LP-relaxations (1.4) suffer from several serious theoretical and practical drawbacks. For instance, it has been shown in [10, 12] that the LP-relaxations *cannot* be exact for most convex problems, i.e., the sequence of the associated optimal values converges to the global optimum only *asymptotically* and not in finitely many steps. Moreover, the LPs of the hierarchy are numerically ill-conditioned. This is in contrast with the semidefinite relaxations (1.3) for which finite convergence takes place for convex problems where $\nabla^2 f(\mathbf{x}^*)$ is positive definite at every minimizer $\mathbf{x}^* \in \mathbf{K}$ (see de Klerk and Laurent [6, Corollary 3.3]) and occurs at the first relaxation for SOS-convex[1] problems [11, Theorem 3.3]. In fact, as demonstrated in recent works of Marshall [14] and Nie [15], finite convergence is generic (even for non convex problems).

So *would it be possible to define a hierarchy of convex relaxations in between (1.3) and (1.4), i.e., with some of the nice features of the semidefinite relaxations but with a much less demanding computational effort (hence closer to the LP-relaxations)*? This paper is a contribution in this direction.

Contribution. This paper consists of two contributions: In the first

[1] An SOS-convex polynomial is a convex polynomial whose Hessian factors as $L(\mathbf{x})L(\mathbf{x})^T$ for some rectangular matrix polynomial L. For instance, separable convex polynomials are SOS-convex.

contribution which is of theoretical nature, we describe a new hierarchy of convex relaxations for $\mathbf{P}$ with the following feature. Each relaxation in the hierarchy is a finite-dimensional convex optimization problem of the form:

$$\rho_d \;=\; \max_{\lambda}\{\, G_d(\lambda) \;:\; \lambda \geq 0\}, \tag{1.5}$$

where $G_d(\cdot)$ is the concave function defined by:

$$G_d(\lambda) \;:=\; \min_{\mathbf{x}}\; \{f(\mathbf{x}) \quad - \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_d} \lambda_{\alpha\beta} \left(\prod_{j=1}^{m} g_j^{\alpha_j}(\mathbf{x})\right) \times \left(\prod_{j=1}^{m}(1-g_j(\mathbf{x}))^{\beta_j}\right) \right\}. \tag{1.6}$$

Therefore $\rho_d \leq f^*$ for all d. And we prove that:

(a) $\rho_d \geq \theta_d$ for all d, and so $\rho_d \to f^*$ as one let d increase.

(b) For convex problems $\mathbf{P}$, i.e., when $f, -g_j$ are convex, $j = 1,\ldots,m$, and Slater's condition holds, the convergence is finite and occurs at the first relaxation, i.e., $\rho_1 = f^*$, in contrast with the LP-relaxations (1.4) where convergence cannot be finite (and is very slow on simple trivial examples). In fact computing ρ_1 is just applying the standard dual method of multipliers (or Lagrangian relaxation) to the convex problem $\mathbf{P}$.

(c) For 0/1 optimization, i.e., when $\mathbf{K} \subseteq \{0,1\}^n$, finite convergence takes place and the optimal value ρ_d provides a better lower bound than the one obtained with Sherali-Adams' RLT hierarchy [17]. In fact, the latter is solving (1.4) with only a subset of the products that appear in (1.4).

(d) Finally, (1.5) has a nice interpretation in terms of the dual method of Non Linear Programming (or Lagrangian relaxation). To see this, consider the optimization problem $\tilde{\mathbf{P}}_d$ defined by:

$$\min_{\mathbf{x}}\,\{f(\mathbf{x}) \;:\; g_j(\mathbf{x})^{\alpha_j}(1-g_j(\mathbf{x}))^{\beta_j} \;\geq\; 0,\; (\alpha,\beta)\in\mathbb{N}^{2m}_d\} \tag{1.7}$$

which has same value f^* as $\mathbf{P}$ because $\tilde{\mathbf{P}}_d$ is just $\mathbf{P}$ with additional redundant constraints; and notice that $\tilde{\mathbf{P}}_1 = \mathbf{P}$. Then solving (1.5) is just applying the *dual method* of multipliers in Non Linear Programming to $\tilde{\mathbf{P}}_d$; see e.g. [4, Chapter 8]. In general one obtains only a lower bound on the optimal value of $\tilde{\mathbf{P}}_d$ when $\mathbf{P}$ is not a convex program). And so our result states that the Lagrangian relaxation applied to $\tilde{\mathbf{P}}_d$ provides a lower bound as close to f^*

as desired, provided that d is sufficiently large, i.e., *provided* that sufficiently many redundant constraints are added to the description of $\mathbf{P}$.

Note in passing that this provides a rigorous *rationale* for the well-known fact that adding redundant constraints helps for solving $\mathbf{P}$. Indeed, even though the new problems $\tilde{\mathbf{P}}_d$, $d \in \mathbb{N}$, are all equivalent to $\mathbf{P}$, their Lagrangian relaxations are *not* equivalent to that of $\mathbf{P}$.

Practical and computational considerations

Our second contribution has a practical and algorithmic flavor. Even though (1.5) is a convex optimization problem, evaluating $G_d(\lambda)$ at a point $\lambda \geq 0$ requires computing the unconstrained global minimum of the function

$$\mathbf{x} \mapsto L_d(\mathbf{x}, \lambda) \;:=\; f(\mathbf{x}) - \sum_{(\alpha,\beta)\in\mathbb{N}^{2m}_d} \lambda_{\alpha\beta} \left(\prod_{j=1}^{m} g_j(\mathbf{x})^{\alpha_j} \right) \times \left(\prod_{j=1}^{m} (1 - g_j(\mathbf{x}))^{\beta_j} \right), \tag{1.8}$$

an NP-hard problem in general. After all, in principle the goal of Lagrangian relaxation is to end up with a problem which is easier to solve than $\mathbf{P}$, and so, in this respect, the hierarchy (1.5) is *not* practical.

So in this second part of the paper, we first show that the LP-relaxations (1.4) can be interpreted as a way to "restrict" and simplify the hierarchy (1.5) by a simple and brute force trick, so as to make it tractable (but of course less efficient). Namely, a certain nonnegative polynomial (whose coefficients have to be determined) is imposed to be the constant polynomial equal to zero! More precisely, the nonnegative vector λ in (1.5) is restricted to a polytope so as to make the polynomial L_d in (1.8) constant! In fact, if one had initially defined the LP-relaxations (1.4) as this brute force (and even brutal) simplification of (1.5), it would have been hard to justify.

Inspired by this interpretation, we propose a systematic way to define improved versions of the LP-hierarchy (1.4) by simplifying (1.5) in a much less brutal manner. We now impose the same nonnegative polynomial $L_d - t$ to be an SOS polynomial of fixed degree $2k$ (rather than the zero polynomial in (1.4)). The increase of complexity is completely controlled by the parameter $k \in \mathbb{N}$ and is chosen by the user. That is, in the new resulting hierarchy (parametrized by k), each LP of the hierarchy (1.4) now becomes a semidefinite program but whose size of the semidefiniteness constraint is fixed and equal to $\binom{n+k}{n}$, independently of the rank d in the hierarchy. (It is

known that crucial for solving semidefinite programs is the size of the LMIs involved rather than the number of variables.) The level $k = 0$ of complexity corresponds to the original LP-relaxations (1.4), the level $k = 1$ corresponds to a hierarchy of semidefinite programs with an Linear Matrix Inequality (LMI) of size $(n + 1)$, etc. To fix ideas, let us mention that for $k = 1$, the first relaxation (i.e., $d = 1$) is even stronger than the first relaxation of the hierarchy (1.3) as it takes into account products of linear constraints; and so for instance, when applied to the celebrated MAXCUT problem, the first relaxation has the Goemans-Williamson's performance guarantee. Moereover, when $k = 1$ one obtains the so-called "Sherali-Adams + SDP" hierarchy already used for approximating some 0/1 optimization problems.

So an important issue is: *What do we gain by this increase of complexity?*

Of course, from a computational complexity point of view, one way got evaluate the efficiency of those relaxations is to analyze whether they help reduce integrality gaps, e.g. for some 0/1 optimization problems. For the level $k = 1$ (i.e. the "Adams-Sherali + SDP hierarchy") some negative results in this direction have been provided in Benabbas and Magen [2], and in Benabbas et al. [3].

But in a different point of view, we claim that a highly desirable property for a general purpose method (e.g., the hierarchies (1.3) or (1.4)) aiming at solving NP-hard optimization problems, is to behave "efficiently" when applied to a class of problems considered relatively "easy" to solve. Otherwise one might raise reasonable doubts on its efficiency for more difficult problems, not only in a worst-case sense but also in "average". Convex problems $\mathbf{P}$ as in (1.1)-(1.2), i.e., when $f, -g_j$ are convex, form the most natural class of problems which are considered easy to solve by some standard methods of Non Linear Programming; see e.g. Ben-tal and Nemirovski [1]. We have already proved that the hierarchy (1.3) somehow recognizes convexity. For instance, finite convergence takes places as soon as $\nabla^2 f(\mathbf{x}^*)$ is positive definite at every global minimizer $\mathbf{x}^* \in \mathbf{K}$ (see deKlerk and Laurent [6]); moreover, SOS-convex programs are solved at the first step of the hierarchy as shown in Lasserre [11]. On the other hand, the LP-hierarchy (1.4) behaves poorly on such problems as the convergence cannot be finite; see e.g. Lasserre [12, 10].

We prove that the gain by this (controlled) increase of complexity is precisely to permit finite convergence (and at the first step of the hierarchy) for a non trivial class of convex problems. For instance with $k = 1$ the resulting hierarchy of semidefinite programs solves convex quadratic programs exactly at the first step of the hierarchy. And more generally, for $k > 1$, the

first relaxation is exact for SOS-convex[2] problems of degree at most k. On the other hand, we show that for non convex problems, exactness at some relaxation in the hierarchy still implies restrictive conditions.

2 Main result

2.1 Notation and definitions

Let $\mathbb{R}[\mathbf{x}]$ be the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$. Denote by $\mathbb{R}[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]$ the vector space of polynomials of degree at most d, which forms a vector space of dimension $s(d) = \binom{n+d}{d}$, with e.g., the usual canonical basis $(\mathbf{x}^\alpha)$ of monomials. Also, denote by $\Sigma[\mathbf{x}] \subset \mathbb{R}[\mathbf{x}]$ (resp. $\Sigma[\mathbf{x}]_d \subset \mathbb{R}[\mathbf{x}]_{2d}$) the space of sums of squares (s.o.s.) polynomials (resp. s.o.s. polynomials of degree at most $2d$). If $f \in \mathbb{R}[\mathbf{x}]_d$, write $f(\mathbf{x}) = \sum_{\alpha \in \mathbb{N}^n_d} f_\alpha \mathbf{x}^\alpha$ in the canonical basis and denote by $\mathbf{f} = (f_\alpha) \in \mathbb{R}^{s(d)}$ its vector of coefficients. Finally, let $\mathcal{S}^n$ denote the space of $n \times n$ real symmetric matrices, with inner product $\langle \mathbf{A}, \mathbf{B} \rangle = \operatorname{trace} \mathbf{AB}$, and where the notation $\mathbf{A} \succeq 0$ (resp. $\mathbf{A} \succ 0$) stands for $\mathbf{A}$ is positive semidefinite. With $g_0 := 1$, the quadratic module $Q(g_1, \ldots, g_m) \subset \mathbb{R}[\mathbf{x}]$ generated by polynomials $g_1, \ldots, g_m$, is defined by

$$Q(g_1, \ldots, g_m) := \{\sum_{j=0}^{m} \sigma_j \, g_j \; : \; \sigma_j \in \Sigma[\mathbf{x}]\}.$$

We briefly recall two important theorems by Putinar [16] and Krivine-Stengle [9, 19] respectively, on the representation of polynomials positive on $\mathbf{K}$,

Theorem 2.1 *Let $g_0 = 1$ and $\mathbf{K}$ in (1.2) be compact.*

(a) If the quadratic polynomial $\mathbf{x} \mapsto M - \|\mathbf{x}\|^2$ belongs to $Q(g_1, \ldots, g_m)$ and if $f \in \mathbb{R}[\mathbf{x}]$ is strictly positive on $\mathbf{K}$ then $f \in Q(g_1, \ldots, g_m)$.

(b) Assume that $0 \leq g_j \leq 1$ on $\mathbf{K}$ for every j, and the family $\{g_j, 1 - g_j\}$ generates $\mathbb{R}[\mathbf{x}]$. If f is strictly positive on $\mathbf{K}$ then

$$f = \sum_{\alpha, \beta \in \mathbb{N}^m} c_{\alpha\beta} \prod_j g_j^{\alpha_j} \prod_\ell (1 - g_\ell)^{\beta_\ell},$$

for some finitely many nonnegative scalars $(c_{\alpha\beta})$.

[2] A SOS-convex polynomial is such that its Hessian matrix is SOS, i.e., factors as $L(\mathbf{x})L(\mathbf{x})^T$ for some rectangular matrix polynomial L.

2.2 Main result

With $\mathbf{K}$ as in (1.2) we make the following assumption:

Assumption 1 $\mathbf{K}$ *is compact and* $0 \leq g_j \leq 1$ *on* $\mathbf{K}$ *for all* $j = 1, \ldots, m$. *Moreover, the family of polynomials* $\{g_j, 1 - g_j\}$ *generates the algebra* $\mathbb{R}[\mathbf{x}]$.

Notice that if $\mathbf{K}$ is compact and Assumption 1 does not hold, one may always rescale the variables x_i so as to have $\mathbf{K} \subset [0,1]^n$, and then add redundant constraints $0 \leq x_i \leq 1$ for all $i = 1, \ldots, m$. Then the family $\{g_j, 1 - g_j\}$ (which includes x_j and $1 - x_j$ for all j) generates the algebra $\mathbb{R}[\mathbf{x}]$ and Assumption 1 holds.

With $d \in \mathbb{N}$ and $0 \leq \lambda = (\lambda_{\alpha\beta})$, $(\alpha, \beta) \in \mathbb{N}^{2m}_d$, let $\lambda \mapsto G_d(\lambda)$ be the function defined in (1.6), with associated problem:

$$\rho_d \;=\; \max_{\lambda} \{\, G_d(\lambda) \;:\; \lambda \geq 0\}. \tag{2.1}$$

Observe that $G_d(\lambda) \leq f^*$ for all $\lambda \geq 0$, and computing ρ_d is just solving the Lagrangian relaxation of problem $\tilde{\mathbf{P}}_d$ in (1.7).

Theorem 2.2 *Let* $\mathbf{K}$ *be as in (1.2),* $f \in \mathbb{R}[\mathbf{x}]$, $d \in \mathbb{N}$, *and let Assumption 1 hold. Consider problem (2.1) associated with* $\mathbf{P}$ *and with optimal value* ρ_d. *Then the sequence* (ρ_d), $d \in \mathbb{N}$, *is monotone nondecreasing and* $\rho_d \to f^*$ *as* $d \to \infty$.

Proof. We first prove that $\rho_{d+1} \geq \rho_d$ for all d, so that the sequence (ρ_d), $d \in \mathbb{N}$, is monotone nondecreasing. Let $0 \leq \lambda = (\lambda_{\alpha\beta})$ with $(\alpha, \beta) \in \mathbb{N}^{2m}_d$. Then $0 \leq \tilde{\lambda}$ with $\tilde{\lambda}_{\alpha\beta} = \lambda_{\alpha\beta}$ whenever $(\alpha, \beta) \in \mathbb{N}^{2m}_d$, and $\tilde{\lambda}_{\alpha\beta} = 0$ whenever $|\alpha + \beta| > d$, is such that $G_{d+1}(\tilde{\lambda}) = G_d(\lambda)$ and so $\rho_{d+1} \geq \rho_d$. Next, let $\epsilon > 0$ be fixed, arbitrary. The polynomial $f - f^* + \epsilon$ is positive on $\mathbf{K}$ and therefore, by [19], [12, Theorem 2.23],

$$f - (f^* - \epsilon) \;=\; \sum_{(\alpha,\beta) \in \mathbb{N}^{2m}_d} c^{\epsilon}_{\alpha\beta} \left(\prod_{j=1}^{m} g_j^{\alpha_j} \right) \left(\prod_{j=1}^{m} (1 - g_j)^{\beta_j} \right),$$

for some nonnegative vector of coefficients $\mathbf{c}^{\epsilon} = (c^{\epsilon}_{\alpha\beta})$. Equivalently,

$$f - \sum_{(\alpha,\beta) \in \mathbb{N}^{2m}_d} c^{\epsilon}_{\alpha\beta} \left(\prod_{j=1}^{m} g_j^{\alpha_j} \right) \left(\prod_{j=1}^{m} (1 - g_j)^{\beta_j} \right) \;=\; (f^* - \epsilon).$$

Letting

$$d_\epsilon := \max_{\alpha,\beta} \{|\alpha+\beta| : c^\epsilon_{\alpha\beta} > 0\},$$

we obtain $f^* \geq G_{d_\epsilon}(\mathbf{c}^\epsilon) = f^* - \epsilon$. And so

$$f^* \geq \max_\lambda \{G_{d_\epsilon}(\lambda) : \lambda \geq 0\} \geq f^* - \epsilon.$$

As $\epsilon > 0$ was arbitrary, the desired result follows. □

Corollary 2.1 *Let* $\mathbf{K}$ *be as in (1.2), Assumption (1) hold and let* $\tilde{\mathbf{P}}_d$, $d \in \mathbb{N}$, *be as in (1.7). Then for every* $\epsilon > 0$ *there exists* $d_\epsilon \in \mathbb{N}$ *such that for every* $d \geq d_\epsilon$, *the Lagrangian relaxation of* $\tilde{\mathbf{P}}_d$, *yields a lower bound* $f^* - \epsilon \leq \rho_d \leq f^*$.

This follows from Theorem 2.2 and the fact that computing ρ_d is just solving the Lagrangian relaxation associated with $\tilde{\mathbf{P}}_d$. So the interpretation of Corollary 2.1 is that the Lagrangian relaxation technique in non convex optimization can provide a lower bound as close as desired to the global optimum f^* provided that it is applied to an equivalent formulation of $\mathbf{P}$ that contains sufficiently many redundant constraints which are products of the original ones. It also provides a rigorous rationale for the well-known fact that adding redundant constraints helps solve $\mathbf{P}$. Indeed, even though the new problems $\tilde{\mathbf{P}}_d$, $d \in \mathbb{N}$, are all equivalent to $\mathbf{P}$, their Lagrangian relaxations are not equivalent to that of $\mathbf{P}$.

2.3 Convex programs

In this section, the set $\mathbf{K}$ is not assumed to be compact.

Theorem 2.3 *Let* $\mathbf{K}$ *be as in (1.2) and assume that* f *and* $-g_j$ *are convex,* $j = 1, \ldots, m$. *Moreover, assume that Slater's condition*[3] *holds and* $f^* > -\infty$.

Then the hierarchy of convex relaxations (1.5) has finite convergence at step $d = 1$, *i.e.,* $\rho_1 = f^*$, *and* $\rho_1 = G_1(\lambda^*)$ *for some nonnegative* $\lambda^* \in \mathbb{R}^m$.

[3]Slater's condition holds for $\mathbf{P}$ if there exists $\mathbf{x}_0 \in \mathbf{K}$ such that $g_j(\mathbf{x}_0) > 0$ for every $j = 1, \ldots, m$.

Proof. This is because the dual method applied to $\mathbf{P}$ (i.e. $\tilde{\mathbf{P}}_1$) converges, i.e.,

$$\begin{aligned} f^* &= \max_{\lambda \geq 0} \left\{ \min_{\mathbf{x}} \{ f(\mathbf{x}) - \sum_{j=1}^m \lambda_j \, g_j(\mathbf{x}) \} \right\} \\ &= \max_{\lambda} \{ G_1(\lambda) \, : \, \lambda \geq 0 \} \, = \, \rho_1. \end{aligned}$$

Next, let $\lambda^{(n)}$ be a maximizing sequence, i.e., $G_1(\lambda^{(n)}) \to f^*$ as $n \to \infty$. Since Slater's condition holds (say at some $\mathbf{x}_0 \in \mathbf{K}$), one has

$$G_1(\lambda^{(0)}) \, \leq \, G_1(\lambda^{(n)}) \, \leq \, f(\mathbf{x}_0) - \sum_{j=1}^m \lambda_j^{(n)} \, g_j(\mathbf{x}_0),$$

for all n, and so $\lambda_j^{(n)} \leq (f(\mathbf{x}_0) - G_1(\lambda^{(0)}))/g_j(\mathbf{x}_0)$ for every $j = 1, \ldots, m$, and all $n \geq 1$. So there is a subsequence (n_k), $k \in \mathbb{N}$, and $\lambda^* \in \mathbb{R}^m_+$, such that $\lambda^{(n_k)} \to \lambda^* \geq 0$ as $k \to \infty$. Finally, let $\mathbf{x} \in \mathbb{R}^n$ be fixed, arbitrary. From

$$G_1(\lambda^{(n_k)}) \, \leq \, f(\mathbf{x}) - \sum_{j=1}^m \lambda_j^{(n_k)} \, g_j(\mathbf{x}), \qquad \forall \, k,$$

letting $k \to \infty$ yields

$$f^* \, \leq \, f(\mathbf{x}) - \sum_{j=1}^m \lambda_j^* \, g_j(\mathbf{x}).$$

As $\mathbf{x} \in \mathbb{R}^n$ was arbitrary, this proves $G_1(\lambda^*) \geq f^*$, which combined with $G_1(\lambda^*) \leq f^*$ yields the desired result $G_1(\lambda^*) = f^*$. □

Observe that this does not hold for the LP-relaxations (1.4) where generically $\theta_d < f^*$ for every $d \in \mathbb{N}$; see e.g. [10, 12].

3 A parametrized hierarchy of semidefinite relaxations

Problem (2.1) is convex but in general the objective function G_d is non differentiable. Moreover, another difficulty is the computation of $G_d(\lambda)$ for each $\lambda \geq 0$ since $G_d(\lambda)$ is the global optimum of the possibly non convex function $(\mathbf{x}, \lambda) \mapsto L_d(\mathbf{x}, \lambda)$ defined in (1.8). So one strategy is to replace (2.1) by a *simpler* convex problem (while preserving the convergence property) as follows.

3.1 Interpreting the LP-relaxations

Observe that the LP-relaxations (1.4) can be written

$$\theta_d \;=\; \max_{\lambda \geq 0, t} \{\; t \;:\; L_d(\mathbf{x},\lambda) - t \;=\; 0, \quad \forall \mathbf{x} \in \mathbb{R}^n \;\}\,, \tag{3.1}$$

where L_d has been defined in (1.8).

And so the LP-relaxations (1.4) can be interpreted as simplifying (2.1) by restricting the nonnegative orthant $\{\lambda : \lambda \geq 0\}$ to its subset of λ's that make the polynomial $\mathbf{x} \mapsto L(\mathbf{x},\lambda) - t$ constant and equal to zero, instead of being only nonnegative. This subset being a polyhedron, solving (3.1) reduces to solving a linear program. At first glance, such an a priori simple and naive brute force simplification might seem unreasonable (to say the least). But of course the LP-relaxations (1.4) where not defined this way. Initially, the Sherali-Adams' RLT hierarchy [17] was introduced for 0/1 programs and finite convergence was proved by using *ad hoc* arguments. But in fact, the rationale behind convergence of the more general LP-relaxations (1.4) is the Krivine-Stengle positivity certificate [12, Theorem 2.23].

However, even though this brute force simplification still preserves the convergence $\theta_d \to f^*$ thanks to [12, Theorem 2.23], we have already mentioned that it also implies serious theoretical (and practical) drawbacks for the resulting LP-relaxations (like slow asymptotic convergence for convex problems and numerical ill-conditioning).

3.2 A parametrized hierarchy of semidefinite relaxations

However, inspired by this interpretation we propose a systematic way to improve the LP-relaxations (1.4) along the same lines but by doing a much less brutal simplification of (2.1). Indeed, one may now impose on the same nonnegative polynomial $\mathbf{x} \mapsto L(\mathbf{x},\lambda) - t$ to be a sum of squares (SOS) polynomial σ of degree at most $2k$ (instead of being constant and equal to zero as in (3.1)), and solve the resulting hierarchy of optimization problems:

$$\begin{array}{rcll} q_d^k \;=\; \max\limits_{\lambda,t,\sigma} & t & & \\ \text{s.t.} & L_d(\mathbf{x},\lambda) - t = \sigma, & \forall \mathbf{x} \in \mathbb{R}^n\;, \\ & \lambda \geq 0, \quad \sigma \in \Sigma[\mathbf{x}]_k & \end{array} \tag{3.2}$$

with $d = 1, 2, \ldots$, and parametrized by k, fixed. (Recall that $\Sigma[\mathbf{x}]_k$ denotes the set of SOS polynomials of degree at most $2k$.) To see that (3.2) is a semidefinite program, write

$$\mathbf{x} \mapsto L_d(\mathbf{x},\lambda) - t \;:=\; \sum_{\beta \in \mathbb{N}^n_s} L_\beta(\lambda, t)\, \mathbf{x}^\beta,$$

where $s = d \max_j[\deg g_j]$ and $L_\beta(\lambda, t)$ is linear in (λ, t) for each $\beta \in \mathbb{N}^n_s$.

Next, for $k \in \mathbb{N}$ such that $2k \leq s$, let $\mathbf{v}_k(\mathbf{x})$ be the vector of the monomial basis $(\mathbf{x}^\beta)$, $\beta \in \mathbb{N}^n_k$, of $\mathbb{R}[\mathbf{x}]_k$, and write

$$\mathbf{v}_k(\mathbf{x})\,\mathbf{v}_k(\mathbf{x})^T \;=\; \sum_{\beta \in \mathbb{N}^n_{2k}} \mathbf{x}^\beta\,\mathbf{B}_\beta,$$

for some appropriate real symmetric matrices $(\mathbf{B}_\beta)$, $\beta \in \mathbb{N}^n_{2k}$. Then problem (3.2) is the semidefinite program:

$$\begin{array}{rcl} q^k_d \;=\; \max\limits_{\lambda,t,\mathbf{Q}} & t & \\ \text{s.t.} & L_\beta(\lambda,t) \;=\; \langle \mathbf{B}_\beta, \mathbf{Q}\rangle, & \forall \beta \in \mathbb{N}^n_{2k} \\ & L_\beta(\lambda,t) \;=\; 0, & \forall \beta \in \mathbb{N}^n_s,\ |\beta| > 2k \\ & \lambda \;\geq\; 0;\ \mathbf{Q} = \mathbf{Q}^T \succeq 0, & \end{array} \tag{3.3}$$

where $\mathbf{Q}$ is a $\binom{n+k}{n} \times \binom{n+k}{n}$ real symmetric matrix.

Of course $q^k_d \geq \theta_d\,(= q^0_d)$ for all d because with $\sigma = 0$ one retrieves (1.4). Moreover in the semidefinite program (3.3), the semidefinite constraint $\mathbf{Q} \succeq 0$ is concerned with a real symmetric $\binom{n+k}{n} \times \binom{n+k}{n}$ matrix, independently of the rank d in the hierarchy. For instance if $k = 1$ then σ is a quadratic SOS and $\mathbf{Q}$ has size $(n+1) \times (n+1)$. In other words, even if the number of variables $\lambda = (\lambda_{\alpha\beta})$ increases fast with d, the LMI constraint $\mathbf{Q} \succeq 0$ has fixed size, in contrast with the semidefinite relaxations (1.3) where the size of the LMIs increases with d. And it is a well-known fact that crucial for solving semidefinite program is the size of the LMIs involved rather than the number of variables.

3.3 Sherali-Adams' RLT for 0/1 programs

Consider 0/1 programs with $f \in \mathbb{R}[\mathbf{x}]$, and feasible set $\mathbf{K} = \{\mathbf{x} : \mathbf{A}\mathbf{x} \leq \mathbf{b}\} \cap \{0,1\}^n$, for some real matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$ and some vector $\mathbf{b} \in \mathbb{R}^m$. The Sherali-Adams's RLT hierarchy [17] belongs to the family of LP-relaxations (1.4) but with a more specific form since $\mathbf{K} \subset [0,1]^n$. Notice that the family $\{1, x_1, (1-x_1), \ldots, x_n, (1-x_n)\}$ generates the algebra $\mathbb{R}[\mathbf{x}]$. Let $g_\ell(\mathbf{x}) = (b - \mathbf{A}\mathbf{x})_\ell$, $\ell = 1, \ldots, m$, and $g_0(\mathbf{x}) = 1$.

Following the definition of the Sherali-Adams' RLT in [17], the resulting linear program at step d in the hierarchy reads:

$$\theta_d = \max_{\lambda \geq 0, t, h} \left\{ t \;:\; f(\mathbf{x}) - t \;=\; \sum_{i=1}^n h_i(\mathbf{x})\, x_i(1-x_i) \right.$$

$$+\sum_{\ell=0}^{m}\sum_{\substack{I,J\subset\{1,\ldots,n\}\\ I\cap J=\emptyset;|I\cup J|\le d}}\lambda_{IJ}^{\ell}\,g_\ell(\mathbf{x})\prod_{i\in I}x_i\prod_{j\in J}(1-x_j);$$

$$h_i\in\mathbb{R}[\mathbf{x}]_{d-1}\quad i=1,\ldots,n\Big\}\,, \tag{3.4}$$

where λ is the nonnegative vector (λ_{IJ}^{ℓ}). (If there are linear equality constraints $g_\ell(\mathbf{x})=0$ the corresponding variables λ_{IJ}^{ℓ} are not required to be nonnegative.) So all products between the g_ℓ's are ignored (see the paragraph before Lemma 1 in [17, p. 414]) even though they might help tighten the relaxations. In the literature the dual LP of (3.4) is described rather than (3.4) itself.

In this context, the problem $\tilde{\mathbf{P}}_d$ equivalent to $\mathbf{P}$ and defined in (1.7) by adding redundant constraints formed with products of original ones, reads:

$$\begin{aligned}\min\{f(\mathbf{x})\ :\ & \mathbf{x}^\alpha\,x_j(1-x_j)\ =\ 0,\ j=1,\ldots,n;\ \alpha\in\mathbb{N}^n_{d-1};\\ & g_\ell(\mathbf{x})\prod_{i\in I}x_j\prod_{j\in J}(1-x_j)\ge 0,\quad \ell=0,\ldots,m,\\ & I,J\subset\{1,\ldots,n\};\ I\cap J=\emptyset;\ |I\cup J|\le d\}.\end{aligned}$$

Hence the 0/1 analogue of (3.2) reads

$$q_d^k=\max_{\lambda\ge 0,t,h}\left\{t\ :\ f(\mathbf{x})-t\ =\ \sigma(\mathbf{x})+\sum_{i=1}^{n}h_i(\mathbf{x})\,x_i(1-x_i)\right.$$

$$+\sum_{\ell=0}^{m}\sum_{\substack{I,J\subset\{1,\ldots,n\}\\ I\cap J=\emptyset;|I\cup J|\le d}}\lambda_{IJ}^{\ell}\,g_\ell(\mathbf{x})\prod_{i\in I}x_i\prod_{j\in J}(1-x_j);$$

$$\left.\sigma\in\Sigma[\mathbf{x}]_k;\quad h_i\in\mathbb{R}[\mathbf{x}]_{d-1}\quad i=1,\ldots,n\right\}. \tag{3.5}$$

For 0/1 programs with linear or quadratic objective function, and for every $k\ge 1$, the first semidefinite relaxation (3.5), i.e., with $d=2$, is at least as powerful as that of the standard hierarchy of semidefinite relaxations (1.3). Indeed (3.5) contains products $g_\ell(\mathbf{x})x_j$ or $g_\ell(\mathbf{x})(1-x_k)$, for all (ℓ,j,k), which do to not appear in (1.3) with $d=1$. And so in particular, the first such relaxation for MAXCUT has the celebrated Goemans-Williamson's performance guarantee while the standard LP-relaxations (1.4) do not. On the other hand, for 0/1 problems and for the parameter value $k=1$, the hierarchy (3.5) is what is called the *Sherali-Adams + SDP* hierarchy (basic SDP-relaxation + RLT hierarchy) in e.g. Benabas and Magen [3] and Benabbas et al. [2]; and in [3, 2] the authors show that any (constant) level d of

this hierarchy, viewed as a strengthening of the basic SDP-relaxation, does not make the integrality gap decrease.

In fact, and in view of our previous analysis, the "Sherali-Adams + SDP" hierarchy should be viewed as a *(level $k = 1$)*-strengthening of the basic Sherali-Adams' LP-hierarchy (3.4) rather than a strengthening of the basic SDP relaxation.

4 Comparing with standard LP-relaxations

As asked in introduction:
What do we gain by going from the LP hierarchy (1.4) to the semidefinite hierarchy (3.3) parametrized by k? Some answers are provided below.

4.1 Convex problems

Recall that a highly desirable property for a general purpose method aiming at solving NP-hard optimization problems, is to behave efficiently when applied to a class of problems considered relatively easy to solve. Otherwise one might raise reasonable doubts on its efficiency for more difficult problems not only in a worst-case sense but also in average. And convex problems **P** as in (1.1)-(1.2), i.e., when $f, -g_j$ are convex, form the most natural class of problems which are considered easy to solve by some standard methods of Non Linear Programming.

Theorem 4.1 *With* **P** *as in (1.1)-(1.2) let $f, -g_j$ be convex, $j = 1, \ldots, m$, let Slater's condition hold and let $f^* > -\infty$. Then:*

(a) If $\max[\deg f, \deg g_j] \leq 2$ then $q_1^1 = f^*$, i.e., the first relaxation of the hierarchy (3.2) parametrized by $k = 1$, is exact.

(a) If $\max[\deg f, \deg g_j] \leq 2k$ and $f, -g_j$ are all SOS-convex, then $q_1^k = f^*$, i.e., the first relaxation of the hierarchy (3.2) parametrized by k, is exact.

Proof. Under the assumptions of Theorem 4.1, **P** has a minimizer $\mathbf{x}^* \in \mathbf{K}$ and the Karush-Kuhn-Tucker optimality conditions hold at $(\mathbf{x}^*, \lambda^*) \in \mathbf{K} \times \mathbb{R}^m_+$ for some $\lambda^* \in \mathbb{R}^m_+$. And so if $k = 1$, the Lagrangian polynomial $L_1(\cdot, \lambda^*) - f^*$ is a nonnegative quadratic polynomial and so an SOS $\sigma^* \in \Sigma[\mathbf{x}]_1$. Therefore as $q_d^1 \leq f^*$ for all d, the triplet $(\lambda^*, f^*, \sigma^*)$ is an optimal solution of (3.2) with $k = d = 1$, which proves (a).

Next, if $k > 1$ and $f, -g_j$ are all SOS-convex then so is the Lagrangian polynomial $L_1(\cdot, \lambda^*) - f^*$. In addition, as $\nabla_{\mathbf{x}} L_1(\mathbf{x}^*, \lambda^*) = 0$ and $L_1(\mathbf{x}^*, \lambda^*) -$

$f^* = 0$, the polynomial $L_1(\cdot, \lambda^*) - f^*$ is SOS; see e.g. Helton and Nie [8, Lemma 4.2]. Hence $L_1(\cdot, \lambda^*) - f^* = \sigma^*$ for some $\sigma^* \in \Sigma[\mathbf{x}]_k$, and again, the triplet $(\lambda^*, f^*, \sigma^*)$ is an optimal solution of (3.2) with $d = 1$, which proves (b). $\square$

Hence by simplifying (1.5) in a less brutal manner than in (1.4) one recovers a nice and highly desirable property for the resulting hierarchy. The price to pay is to pass from solving a hierarchy of LPs to solving hierarchy of semidefinite programs; however the increase in complexity is controlled by the parameter k since the size of the LMI in the semidefinite program (3.3) is $O(n^k)$, independently of the rank d in the hierarchy.

4.2 Obstructions to Exactness

On the other hand, for non convex problems, exactness at level-d of the hierarchy (3.2), i.e., finite convergence after d rounds, still implies restrictive conditions on the problem:

Corollary 4.1 *Let* $\mathbf{P}$ *be as in (1.1)-(1.2) and let Assumption 1 hold. Let* $\mathbf{x}^* \in \mathbf{K}$ *be a global minimizer and let* $I_1(\mathbf{x}^*) := \{j \in \{1, \ldots, m\} : g_j(\mathbf{x}^*) = 0\}$ *and* $I_2(\mathbf{x}^*) := \{j \in \{1, \ldots, m\} : (1 - g_j(\mathbf{x}^*)) = 0\}$ *be the set of active constraints at* $\mathbf{x}^*$*. Let* $0 \leq k \in \mathbb{N}$ *be fixed.*

The level-d semidefinite relaxation (3.2) is exact only if f^* *(resp.* $\mathbf{x}^* \in \mathbf{K}$*) is also the global optimum (resp. a global minimizer) for the problem*

$$\min_{\mathbf{x}} \{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{V}\}, \tag{4.1}$$

where $\mathbf{V} \subset \mathbb{R}^n$ *(see (4.2) below) is a variety defined from some products of the polynomials* g_j*'s and* $(1 - g_j)$*'s. And if* $k = 0$ *then* f *must be constant on the variety* $\mathbf{V}$*.*

Proof. If (3.2) is exact at level $d \in \mathbb{N}$, then

$$L_d(\mathbf{x}, \lambda^*) - f^* = \sigma(\mathbf{x}), \qquad \forall \mathbf{x} \in \mathbb{R}^n,$$

for some $\lambda^* \geq 0$ and some $\sigma \in \Sigma[\mathbf{x}]_k$. Equivalently,

$$f(\mathbf{x}) - f^* \,=\, \sigma(\mathbf{x}) \quad + \sum_{(\alpha,\beta) \in \mathbb{N}^{2m}_d} \lambda^*_{\alpha\beta} \left(\prod_{j=1}^m g_j(\mathbf{x})^{\alpha_j} \right) \times \left(\prod_{j=1}^m (1 - g_j(\mathbf{x}))^{\beta_j} \right).$$

Then evaluating at $\mathbf{x} = \mathbf{x}^*$ yields $\sigma(\mathbf{x}^*) = 0$ and

$$\lambda^*_{\alpha\beta} > 0 \Rightarrow \quad \begin{cases} \exists j \in I_1(\mathbf{x}^*) \text{ s.t. } \alpha_j > 0, \text{ or} \\ \exists j \in I_2(\mathbf{x}^*) \text{ s.t. } \beta_j > 0. \end{cases}$$

So let $\Omega := \{(\alpha,\beta) \in \mathbb{N}^{2m}_d \,:\, \lambda^*_{\alpha\beta} > 0\}$ and for every $(\alpha,\beta) \in \Omega$ let

$$\begin{aligned} J^1_{\alpha\beta} &:= \{j \in I_1(\mathbf{x}^*) \,:\, \alpha_j > 0\}, \\ J^2_{\alpha\beta} &:= \{j \in I_2(\mathbf{x}^*) \,:\, \beta_j > 0\}. \end{aligned}$$

Next, define $\mathbf{V} \subset \mathbb{R}^n$ to be the real variety:

$$\begin{aligned} \{\mathbf{x} \in \mathbb{R}^n \;:\; & \left(\prod_{j \in J^1_{\alpha\beta}} g_j(\mathbf{x})\right)\left(\prod_{j \in J^2_{\alpha\beta}} (1 - g_j(\mathbf{x}))\right) = 0, \\ & \forall\, (\alpha,\beta) \in \Omega \,\}\,. \end{aligned} \tag{4.2}$$

Then for every $\mathbf{x} \in \mathbf{V}$, one obtains $f(\mathbf{x}) - f^* = \sigma(\mathbf{x}) \geq 0$, which means that f^* is the global minimum of f on $\mathbf{V}$. If $k = 0$ then σ is constant and equal to zero. And so $f(\mathbf{x}) - f^* = 0$ for all $\mathbf{x} \in \mathbf{V}$. □

Hence Corollary 4.1 shows that exactness at some step d of the hierarchy (3.2) imposes rather restrictive conditions on problem **P**. Namely, the global optimum f^* (resp. the global minimizer $\mathbf{x}^* \in \mathbf{K}$) must also be the global optimum (resp. a global minimizer) of problem (4.1). For instance, suppose that only one constraint, say $g_k(\mathbf{x}) \geq 0$, is active at $\mathbf{x}^*$. Then f^* (resp. $\mathbf{x}^*$) is also the global minimum (resp. a global minimizer) of f on the variety $\{\mathbf{x} \,:\, g_k(\mathbf{x}) = 0\}$. And if $k = 0$ then f must be constant on the variety $\mathbf{V}$!

Example 1 *If* $\mathbf{K}$ *is the (compact) polytope* $\{\mathbf{x} \,:\, \mathbf{a}_j^T\mathbf{x} \leq 1,\, j = 1,\ldots,m\}$ *for some vectors* $(\mathbf{a}_j) \subset \mathbb{R}^n$*, then invoking a result by Handelman [7], one does not need the polynomials* $\{1 - g_j\}$ *in the definition (1.8) of* L_d*. So for instance, suppose that* $I_1(\mathbf{x}^*) = \{\ell\}$ *at a global minimizer* $\mathbf{x}^* \in \mathbf{K}$*. Then exactness at some step* d *of the hierarchy (3.2) imposes that* f^* *should also be the global minimum of* f *on the whole hyperplane* $\mathbf{V} = \{\mathbf{x} : \mathbf{a}_\ell^T\mathbf{x} = 1\}$*; for non convex functions* f*, this is a serious restriction. Moreover, if* $k = 0$ *then* f *must be constant on the hyperplane* $\mathbf{V}$*.*

Concerning exactness for 0/1 polynomial optimization:

Corollary 4.2 *Let* $\mathbf{K} = \{\mathbf{x} \,:\, \mathbf{A}\mathbf{x} \leq \mathbf{b}\} \cap \{0,1\}^n$ *and let* $\mathbf{x}^* \in \mathbf{K}$ *be an optimal solution of* $f^* = \min\{f(\mathbf{x}) : \mathbf{x} \in \mathbf{K}\}$*. Assume that* $\mathbf{A}\mathbf{x}^* < \mathbf{b}$*, i.e., no constraint is active at* $\mathbf{x}^*$*.*

(a) The Sherali-Adams' RLT relaxation (3.4) is exact at step d in the hierarchy only if $f(\mathbf{x}) = f(\mathbf{x}^*) = f^*$ *for all* $\mathbf{x}$ *in the set*

$$\mathbf{V} := \{\mathbf{x} \in \{0,1\}^n \,:\, \prod_{i\in I} x_i \prod_{j\in J}(1-x_j) = 0, \quad (I,J) \in Q\},$$

where Q *is some finite set of couples* (I,J) *satisfying* $I \cap J = \emptyset$ *and* $|I \cup J| \le d$.

(b) Similarly, the semidefinite relaxation (3.5) is exact at step d only if $\mathbf{x}^*$ *is also a global minimizer of* $\min\{f(\mathbf{x}) \,:\, \mathbf{x} \in \mathbf{V}\}$ *for some* $\mathbf{V}$ *as in (a).*

Proof. (a) Exactness implies that the polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - f^*$ has the representation described in (3.4) for some polynomials $(h_i) \subset \mathbb{R}[\mathbf{x}]_{d-2}$ and some nonnegative scalars (λ^ℓ_{IJ}). Evaluating both sides of (3.4) at $\mathbf{x} = \mathbf{x}^*$ and using $g_\ell(\mathbf{x}^*) > 0$ for all $\ell = 0, \ldots, m$, yields

$$\lambda^\ell_{IJ} > 0 \quad \Longrightarrow \quad \prod_{i\in I} x^*_i \prod_{j\in J}(1-x^*_j) = 0. \tag{4.3}$$

Let $\mathbf{V}$ be as in Corollary 4.2 with $Q := \{(I,J) \,:\, \exists\, \ell \text{ s.t. } \lambda^\ell_{IJ} > 0\}$. Then from the representation of $\mathbf{x} \mapsto f(\mathbf{x}) - f^*$ in (3.4) we obtain $\mathbf{f}(\mathbf{x}) - f^* = 0$ for all $\mathbf{x} \in \mathbf{V}$ and the result follows. For (b) a similar argument is valid but now using the representation of $f(\mathbf{x}) - f^*$ described in (3.5). And so exactness yields (4.3) as well as $\sigma(\mathbf{x}^*) = 0$. Next, for every $\mathbf{x} \in \mathbf{V}$ we now obtain $f(\mathbf{x}) - f^* = \sigma(\mathbf{x}) \ge 0$ because σ is SOS. □

The constraints $\mathbf{A}\mathbf{x} \le \mathbf{b}$ play no explicit role in the definition of the set $\mathbf{V}$. Moreover, if f discriminates all points of the hypercube $\{0,1\}^n$ then exactness of the Sherali-Adams' RLT implies that $\mathbf{V}$ must be the singleton $\{\mathbf{x}^*\}$.

On the hierarchy of semidefinite relaxations

Similarly, the hierarchy of semidefinite relaxations (1.3) also has an interpretation in terms of simplifying an *extended* Lagrangian relaxation of $\mathbf{P}$. Indeed consider the hierarchy of optimization problems

$$\begin{aligned}\omega_d := \max_{\sigma_j} \{H(\sigma_1, \ldots, \sigma_m) \;:\; & \deg(\sigma_j\, g_j) \le 2d, \sigma_j \in \Sigma[\mathbf{x}] \\ & j = 1, \ldots, m\},\end{aligned} \tag{4.4}$$

$d \in \mathbb{N}$, where $\sigma \mapsto H(\sigma_1, \ldots, \sigma_m)$ is the function

$$H(\sigma_1, \ldots, \sigma_m) := \min_{\mathbf{x}} \{f(\mathbf{x}) - \sum_{j=1}^m \sigma_j(\mathbf{x})\, g_j(\mathbf{x})\}.$$

For each $d \in \mathbb{N}$, problem (4.4) is an obvious relaxation of $\mathbf{P}$ and in fact is an *extended Lagrangian relaxation* of $\mathbf{P}$ where the multipliers are now allowed to be SOS polynomials with a degree bound, instead of constant nonnegative polynomials (i.e., SOS polynomials of degree zero).

If $\mathbf{K}$ is compact and the quadratic module

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

(where $g_0 = 1$) is Archimedean, then $\omega_d \to f^*$ as $d \to \infty$. But of course, and like for the usual Lagrangian, minimizing the extended Lagrangian

$$\mathbf{x} \mapsto L(\mathbf{x}, \sigma) := f(\mathbf{x}) - \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}),$$

is in general an NP-hard problem. In fact, writing (4.4) as

$$\begin{aligned}\omega_d = \max_{t,\sigma}\{t \quad : \quad & f(\mathbf{x}) - \sum_{j=1}^{m} \sigma_j(\mathbf{x})\, g_j(\mathbf{x}) - t \;\geq\; 0 \quad \forall \mathbf{x}\,; \\ & \deg(\sigma_j g_j) \leq 2d\},\end{aligned}$$

the semidefinite relaxations (1.3) simplify (4.4) by imposing on the nonnegative polynomial $\mathbf{x} \mapsto f(\mathbf{x}) - \sum_j \sigma_j(\mathbf{x}) g_j(\mathbf{x}) - t$ to be an SOS polynomial $\sigma_0 \in \Sigma[\mathbf{x}]_d$ (rather than just being nonnegative).

But the spirit is different from the LP-relaxations as there is no problem $\tilde{\mathbf{P}}_d$ obtained from $\mathbf{P}$ by adding finitely many redundant constraints and equivalent to $\mathbf{P}$. Instead of adding more and more redundant constraints and doing a standard Lagrangian relaxation to $\tilde{\mathbf{P}}_d$, one applies an extended Lagrangian relaxation to $\mathbf{P}$ with SOS multipliers of increasing degree (instead of nonnegative scalars). And in contrast to LP-relaxations, there is no obstruction to exactness (i.e., finite convergence). In fact, it is quite the opposite since as demonstrated recently in Nie [15], finite convergence is generic!

5 Conclusion

We have shown that the hierarchy of LP-relaxations (1.4) has a rather surprising interpretation in terms of the Lagrangian relaxation applied to a problem $\tilde{\mathbf{P}}$ equivalent to $\mathbf{P}$ (but with redundant constraints formed with

product of polynomials defining the original constraints of $\mathbf{P}$). Indeed it consists of the brute force simplification of imposing on a certain nonnegative polynomial to be the constant polynomial equal to zero, a very restrictive condition.

However, inspired by this interpretation, one has provided a systematic strategy to improve the LP-hierarchy by doing a much less brutal simplification. That is, one now imposes on the same nonnegative polynomial to be an SOS polynomial whose degree k is fixed in advance and parametrizes the whole hierarchy. Each convex relaxation is now a semidefinite program but whose LMI constraint has fixed size $O(n^k)$. Hence, the resulting families of parametrized relaxations achieve a compromise between the hierarchy of semidefinite relaxations (1.3) limited to problems of modest size and the LP-relaxations (1.4) that theoretically can handle problems of larger size but with a poor behavior when applied to convex problems.

References

[1] Ben-Tal A., Nemirovski A.. *Lectures on Modern Convex Optimization*, SIAM, Philadelphia, 2001.

[2] Benabbas B., Georgiou K., Magen A., Tulsiani M. SDP gaps from pairwise independence, Theory of Computing **8**, pp. 269–289, 2012.

[3] Benabbas S., Magen A. Extending SDP integrality gaps to Sherali-Adams with applications to Quadratic Programming and MaxCut-Gain, in *Integer Programming and Combinatorial Optimization*, Lecture Notes in Computer Science, Springer 2010, pp. 299–312.

[4] Bertsekas D.P., Nedić A., Ozdaglar E. *Convex Analysis and Optimization*, Athena Scientific, Belmont, Massachusetts, 2003.

[5] Chlamtac E., Tulsiani M. Convex relaxations and integrality gaps, in *Handbook of Semidefinite, Conic and Polynomial Optimization*, M. Anjos and J.B. Lasserre Eds., Springer, New York, 2012, pp. 139–170.

[6] de Klerk E., Laurent M. On the Lasserre hierarchy of semidefinite programming relaxations of convex polynomial optimization problems, SIAM J. Optim. **21**, pp. 824–832, 2011.

[7] Handelman D. Representing polynomials by positive linear functions on compact convex polyhedra, Pac. J. Math. **132**, pp. 35–62, 1988.

[8] Helton J.W., Nie J. Semidefinite representation of convex sets and convex hulls, in *Handbook on Semidefinite, Conic and Polynomial Optimization*, M. Anjos and J.B. Lasserre Eds., Springer, New York, 2012, pp. 77–112.

[9] Krivine J.L. Anneaux préordonnés, J. Anal. Math. **12**, pp. 307–326, 1964.

[10] Lasserre J.B. Semidefinite programming vs. LP relaxations for polynomial programming, Math. Oper. Res. **27**, pp. 347–360, 2002.

[11] Lasserre J.B. Convexity in semi-algebraic geometry and polynomial optimization, SIAM J. Optim. **19**, pp. 1995–2014, 2009.

[12] Lasserre J.B. *Moments, Positive Polynomials and Their Applications*, Imperial College Press, London, 2009.

[13] Laurent M. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming, Math. Oper. Res. **28**, pp. 470–496, 2003.

[14] Marshall M. Representation of non-negative polynomials, degree bounds and applications to optimization, Canad. J. Math. **61**, pp. 205–221, 2009.

[15] Nie J. Optimality Conditions and Finite Convergence of Lasserre's Hierarchy, Technical report, Dept. Mathematics, University of California, San Diego, 2012. `arXiv:1206.0319`.

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

[17] Sherali H.D., Adams W.P. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems, SIAM J. Discr. Math. **3**, pp. 411–430, 1990.

[18] Sherali H.D., Adams W.P. *A Reformulation-Linearization Technique for Solving Discrete and Continuous Nonconvex Problems*, Kluwer, Dordrecht, MA, 1999.

[19] Stengle G. A Nullstellensatz and a Positivstellensatz in semialgebraic geometry, Math. Ann. **207**, pp. 87–97, 1974.