A "joint+marginal" algorithm for polynomial optimization

Jean-Bernard Lasserre, Thanh Tung Phan

A "joint+marginal" algorithm for polynomial optimization

Jean B. Lasserre and Tung Phan Thanh

***Abstract*— We present a new algorithm for solving a polynomial program $\mathbf{P}$ based on the recent "joint + marginal" approach of the first author for parametric polynomial optimization. The idea is to first consider the variable x_1 as a *parameter* and solve the associated $(n-1)$-variable $(x_2,\ldots,x_n)$ problem $\mathbf{P}(x_1)$ where the parameter x_1 is fixed and takes values in some interval $\mathbf{Y}_1 \subset \mathbb{R}$, with some probability φ_1 uniformly distributed on $\mathbf{Y}_1$. Then one considers the hierarchy of what we call "joint+marginal" semidefinite relaxations, whose duals provide a sequence of univariate polynomial approximations $x_1 \mapsto p_k(x_1)$ that converges to the optimal value function $x_1 \mapsto J(x_1)$ of problem $\mathbf{P}(x_1)$, as k increases. Then with k fixed *à priori*, one computes $\tilde{x}_1^* \in \mathbf{Y}_1$ which minimizes the univariate polynomial $p_k(x_1)$ on the interval $\mathbf{Y}_1$, a convex optimization problem that can be solved via a single semidefinite program. The quality of the approximation depends on how large k can be chosen (in general for significant size problems $k=1$ is the only choice). One iterates the procedure with now an $(n-2)$-variable problem $\mathbf{P}(x_2)$ with parameter x_2 in some new interval $\mathbf{Y}_2 \subset \mathbb{R}$, etc. so as to finally obtain a vector $\tilde{\mathbf{x}} \in \mathbb{R}^n$. Preliminary numerical results are provided.**

I. INTRODUCTION

Consider the general polynomial program

$$\mathbf{P}: \quad f^* := \min_{\mathbf{x}} \{ f(\mathbf{x}) \; : \; \mathbf{x} \in \mathbf{K} \} \tag{1}$$

where f is a polynomial, $\mathbf{K} \subset \mathbb{R}^n$ is a basic semi-algebraic set, and f^* is the *global* minimum of $\mathbf{P}$ (as opposed to a local minimum). One way to approximate the global optimum f^* of $\mathbf{P}$ is to solve a hierarchy of either LP-relaxations or semidefinite relaxations as proposed in e.g. Lasserre [4], [5]. Despite practice with the semidefinite relaxations seems to reveals that convergence is fast, the matrix size in the i-th semidefinite relaxation of the hierarchy grows up as fast as $O(n^i)$. Hence, for large size (and sometimes even medium size) problems, only a few relaxations of the hierarchy can be implemented (the first, second or third relaxation). In that case, one only obtains a lower bound on f^*, and no feasible solution in general. So an important issue is:

How can we use the result of the i-th semidefinite relaxation to find an approximate feasible solution of the original problem?

For some well-known special cases of 0/1 optimization like e.g. the celebrated MAXCUT problem, one may generate a feasible solution with guaranteed performance, from a randomized rounding procedure that uses an optimal solution of the first semidefinite relaxation (i.e. with $i=1$); see Goemans and Williamson [2]. But in general there is no such procedure.

Our contribution is to provide two relatively simple algorithms for polynomial programs which builds up upon the so-called "joint+marginal" approach (in short (J+M)) developed in [6] for *parametric* polynomial optimization. The (J+M)-approach for variables $\mathbf{x} \in \mathbb{R}^n$ and parameters $\mathbf{y}$ in a simple set $\mathbf{Y}$, consists of the standard hierarchy of semidefinite relaxations in [4] where one treats the parameters $\mathbf{y}$ also as variables. But now the moment-approach implemented in the semidefinite relaxations, considers a *joint* probability distribution on the pair $(\mathbf{x},\mathbf{y})$, with the additional constraint that the *marginal* distribution on $\mathbf{Y}$ is fixed (e.g. the uniform probability distribution on $\mathbf{Y}$); whence the name *"joint+marginal"*.

For every $k=1,\ldots,n$, let the compact interval $\mathbf{Y}_k := [\underline{x}_k, \overline{x}_k] \subset \mathbb{R}$ be contained in the projection of $\mathbf{K}$ into the x_k-coordinate axis. In the context of the (non-parametric) polynomial optimization (1), the above (J+M)-approach can be used as follows in what we call the **(J+M)-algorithm**:

• (a) Treat x_1 as a parameter in the compact interval $\mathbf{Y}_1 = [\underline{x}_1, \overline{x}_1]$ with associated probability distribution φ_1 uniformly distributed on $\mathbf{Y}_1$.

• (b) with $i \in \mathbb{N}$ fixed, solve the i-th semidefinite relaxation of the (J+M)-hierarchy [6] applied to problem $\mathbf{P}(x_1)$ with $n-1$ variables $(x_2,\ldots,x_n)$ and parameter x_1, which is problem $\mathbf{P}$ with the additional constraint that the variable $x_1 \in \mathbf{Y}_1$ is fixed. The dual provides a univariate polynomial $x_1 \mapsto J_i^1(x_1)$ which, if i would increase, would converge to $J^1(x_1)$ in the $L_1(\varphi_1)$-norm. (The map $v \mapsto J^1(v)$ denotes the optimal value function of $\mathbf{P}(v)$, i.e. the optimal value of $\mathbf{P}$ given that the variable x_1 is fixed at the value v.) Next, compute $\tilde{x}_1 \in \mathbf{Y}_1$, a global minimizer of the univariate polynomial J_i^1 on $\mathbf{Y}_1$ (e.g. this can be done by solving a single semidefinite program). Ideally, when i is large enough, $\tilde{x}_1$ should be close to the first coordinate $x_1^*$of a global minimizer $\mathbf{x}^* = x_1^*,\ldots,x_n^*)$ of $\mathbf{P}$.

• (c) go back to step (b) with now $x_2 \in \mathbf{Y}_2 \subset \mathbb{R}$ instead of x_1, and with φ_2 being the probability measure uniformly distributed on $\mathbf{Y}_2$. With the same method, compute a global minimizer $\tilde{x}_2 \in \mathbf{Y}_2$, of the univariate polynomial $x_2 \mapsto J_i^2(x_2)$ on the interval $\mathbf{Y}_2$. Again, if i would increase, J_i^2 would converge in the $L_1(\varphi_2)$-norm to the optimal value function $v \mapsto J^2(v)$ of $\mathbf{P}(x_2)$ (i.e. the optimal value of $\mathbf{P}$ given that the variable x_2 is fixed at the value v.) Iterate until one has obtained $\tilde{x}_n \in \mathbf{Y}_n \subset \mathbb{R}$.

One ends up wih a point $\tilde{\mathbf{x}} \in \prod_{k=1}^n \mathbf{Y}_k$ and in general $\tilde{\mathbf{x}} \notin \mathbf{K}$. One may then use $\tilde{\mathbf{x}}$ as initial guess of a local optimization procedure to find a local minimum $\hat{\mathbf{x}} \in \mathbf{K}$.

J.B. Lasserre is with LAAS-CNRS and the Institute of Mathematics, University of Toulouse, France. lasserre@laas.fr

Tung Phan Thanh is with LAAS-CNRS, University of Toulouse, France. tphanta@laas.fr

The rational behind the (J+M)-algorithm is that if i is large enough and $\mathbf{P}$ has a unique global minimizer $\mathbf{x}^* \in \mathbf{K}$, then $\tilde{\mathbf{x}}$ as well as $\hat{\mathbf{x}}$ should be close to $\mathbf{x}^*$.

The computational complexity before the local optimization procedure is less than solving n times the i-th semidefinite relaxation in the (J+M)-hierarchy (which is itself of same order as the i-th semidefinite relaxation in the hierarchy defined in [4]), i.e., a polynomial in the input size of $\mathbf{P}$.

When the feasible set $\mathbf{K}$ is convex, one may define the following variant to obtain a *feasible* point $\tilde{\mathbf{x}} \in \mathbf{K}$. Again, let $\mathbf{Y}_1$ be the projection of $\mathbf{K}_1$ into the x_1-coordinate axis. Once $\tilde{x}_1 \in \mathbf{Y}_1$ is obtained in step (b), consider the new optimization problem $\mathbf{P}(\tilde{x}_1)$ in the $n-1$ variables $(x_2, \ldots, x_n)$, obtained from $\mathbf{P}$ by fixing the variable $x_1 \in \mathbf{Y}_1$ at the value $\tilde{x}_1$. Its feasible set is the convex set $\mathbf{K}_1 := \mathbf{K} \cap \{\mathbf{x} : x_1 = \tilde{x}_1\}$. Let $\mathbf{Y}_2$ be the projection of $\mathbf{K}_1$ into the x_2-coordinate axis. Then go back to step (b) with now $x_2 \in \mathbf{Y}_2$ as parameter and $(x_3, \ldots, x_n)$ as variables, to obtain a point $\tilde{x}_2 \in \mathbf{Y}_2$, etc. until a point $\tilde{\mathbf{x}} \in \prod_{k=1}^n \mathbf{Y}_k$ is obtained. Notice that now $\tilde{\mathbf{x}} \in \mathbf{K}$ because $\mathbf{K}$ is convex. Then proceed as before with $\tilde{\mathbf{x}}$ being the initial guess of a local minimization algorithm to obtain a local minimizer $\hat{\mathbf{x}} \in \mathbf{K}$ of $\mathbf{P}$.

II. THE "JOINT+MARGINAL APPROACH TO PARAMETRIC OPTIMIZATION

Most of the material of this section is taken from [6]. Let $\mathbb{R}[\mathbf{x}, \mathbf{y}]$ denote the ring of polynomials in the variables $\mathbf{x} = (x_1, \ldots, x_n)$, and the variables $\mathbf{y} = (y_1, \ldots, y_p)$, whereas $\mathbb{R}[\mathbf{x}, \mathbf{y}]_d$ denotes its subspace of polynomials of degree at most d. Let $\Sigma[\mathbf{x}, \mathbf{y}] \subset \mathbb{R}[\mathbf{x}, \mathbf{y}]$ denote the subset of polynomials that are sums of squares (in short s.o.s.). For a real symmetric matrix $\mathbf{A}$ the notation $\mathbf{A} \succeq 0$ stands for $\mathbf{A}$ is positive semidefinite.

The parametric optimization problem

Let $\mathbf{Y} \subset \mathbb{R}^p$ be a compact set, called the *parameter* set, and let $f, h_j \in \mathbb{R}[\mathbf{x}]$, $j = 1, \ldots, m$. Let $\mathbf{K} \subset \mathbb{R}^n \times \mathbb{R}^p$ be the basic closed semi-algberaic set:

$$\mathbf{K} := \{(\mathbf{x}, \mathbf{y}) \,:\, \mathbf{y} \in \mathbf{Y}\,;\, h_j(\mathbf{x}, \mathbf{y}) \,\geq\, 0,\; j = 1, \ldots, m\} \quad (2)$$

and for each $\mathbf{y} \in \mathbf{Y}$, let

$$\mathbf{K}_{\mathbf{y}} \,:=\, \{\, \mathbf{x} \in \mathbb{R}^n \,:\, (\mathbf{x}, \mathbf{y}) \,\in\, \mathbf{K} \,\}. \quad (3)$$

For each $\mathbf{y} \in \mathbf{Y}$, fixed, consider the optimization problem:

$$J(\mathbf{y}) \,:=\, \inf_{\mathbf{x}} \{\, f(\mathbf{x}, \mathbf{y}) \,:\, (\mathbf{x}, \mathbf{y}) \,\in\, \mathbf{K} \,\}. \quad (4)$$

The interpretation is as follows: $\mathbf{Y}$ is a set of parameters and for each instance $\mathbf{y} \in \mathbf{Y}$ of the parameter, one wishes to compute an optimal *decision* vector $\mathbf{x}^*(\mathbf{y})$ that solves problem (4). Let φ be a Borel probability measure on $\mathbf{Y}$, with a positive density with respect to the Lebesgue measure on $\mathbb{R}^p$ (or with respect to the counting measure if $\mathbf{Y}$ is discrete). For instance

$$\varphi(B) \,:=\, \left(\int_{\mathbf{Y}} d\mathbf{y} \right)^{-1} \int_{\mathbf{Y} \cap B} d\mathbf{y}, \qquad \forall B \in \mathcal{B}(\mathbb{R}^p),$$

is uniformly distributed on $\mathbf{Y}$. Sometimes, e.g. in the context of optimization with data uncertainty, φ is already specified. The idea is to use φ (or more precisely, its moments) to get information on the distribution of optimal solutions $\mathbf{x}^*(\mathbf{y})$ of $\mathbf{P}_{\mathbf{y}}$, viewed as random vectors. In this section we assume that for every $\mathbf{y} \in \mathbf{Y}$, the set $\mathbf{K}_{\mathbf{y}}$ in (3) is nonempty.

A. A related infinite-dimensional linear program

Let $\mathbf{M}(\mathbf{K})$ be the set of finite Borel probability measures on $\mathbf{K}$, and consider the following infinite-dimensional linear program $\mathbf{P}$:

$$\rho \,:=\, \inf_{\mu \in \mathbf{M}(\mathbf{K})} \left\{ \int_{\mathbf{K}} f \, d\mu \,:\, \pi\mu \,=\, \varphi \right\}, \quad (5)$$

where $\pi\mu$ denotes the marginal of μ on $\mathbb{R}^p$, that is, $\pi\mu$ is a probability measure on $\mathbb{R}^p$ defined by $\pi\mu(B) := \mu(\mathbb{R}^n \times B)$ for all $B \in \mathcal{B}(\mathbb{R}^p)$. Notice that $\mu(\mathbf{K}) = 1$ for any feasible solution μ of $\mathbf{P}$. Indeed, as φ is a probability measure and $\pi\mu = \varphi$ one has $1 = \varphi(\mathbf{Y}) = \mu(\mathbb{R}^n \times \mathbb{R}^p) = \mu(\mathbf{K})$.

The dual of $\mathbf{P}$ is the the following infinite-dimensional linear program:

$$\begin{array}{rll} \rho^* \,:=\, & \displaystyle\sup_{p \in \mathbb{R}[\mathbf{y}]} & \displaystyle\int_{\mathbf{Y}} p(\mathbf{y}) \, d\varphi(\mathbf{y}) \\ & & f(\mathbf{x}) - p(\mathbf{y}) \,\geq\, 0 \quad \forall (\mathbf{x}, \mathbf{y}) \in \mathbf{K}. \end{array} \quad (6)$$

Recall that a sequence of measurable functions (g_n) on a measure space $(\mathbf{Y}, \mathcal{B}(\mathbf{Y}), \varphi)$ converges to g, *φ-almost uniformly*, if and only if for every $\epsilon > 0$, there is a set $A \in \mathcal{B}(\mathbf{Y})$ such that $\varphi(A) < \epsilon$ and $g_n \to g$, uniformly on $\mathbf{Y} \setminus A$.

Theorem 1 ([6]): Let both $\mathbf{Y} \subset \mathbb{R}^p$ and $\mathbf{K}$ in (2) be compact and assume that for every $\mathbf{y} \in \mathbf{Y}$, the set $\mathbf{K}_{\mathbf{y}} \subset \mathbb{R}^n$ in (3) is nonempty. Let $\mathbf{P}$ be the optimization problem (5) and let $\mathbf{X}^*_{\mathbf{y}} := \{\mathbf{x} \in \mathbb{R}^n \,:\, f(\mathbf{x}, \mathbf{y}) = J(\mathbf{y})\}$, $\mathbf{y} \in \mathbf{Y}$. Then:

(a) $\rho = \displaystyle\int_{\mathbf{Y}} J(\mathbf{y}) \, d\varphi(\mathbf{y})$ and $\mathbf{P}$ has an optimal solution.

(b) Assume that for φ-almost $\mathbf{y} \in \mathbf{Y}$, the set of minimizers of $\mathbf{X}^*_{\mathbf{y}}$ is the singleton $\{\mathbf{x}^*(\mathbf{y})\}$ for some $\mathbf{x}^*(\mathbf{y}) \in \mathbf{K}_{\mathbf{y}}$. Then there is a measurable mapping $g : \mathbf{Y} \to \mathbf{K}_{\mathbf{y}}$ such that

$$\begin{array}{rcl} g(\mathbf{y}) & = & \mathbf{x}^*(\mathbf{y}) \;\text{ for every }\; \mathbf{y} \in \mathbf{Y} \\ \rho & = & \displaystyle\int_{\mathbf{Y}} f(g(\mathbf{y}), \mathbf{y}) \, d\varphi(\mathbf{y}), \end{array} \quad (7)$$

and for every $\alpha \in \mathbb{N}^n$, and $\beta \in \mathbb{N}^p$:

$$\int_{\mathbf{K}} \mathbf{x}^\alpha \mathbf{y}^\beta \, d\mu^*(\mathbf{x}, \mathbf{y}) \,=\, \int_{\mathbf{Y}} \mathbf{y}^\beta \, g(\mathbf{y})^\alpha \, d\varphi(\mathbf{y}). \quad (8)$$

(c) There is no duality gap between (5) and (6), i.e. $\rho = \rho^*$, and if $(p_i)_{i \in \mathbb{N}} \subset \mathbb{R}[\mathbf{y}]$ is a maximizing sequence of (6) then:

$$\int_{\mathbf{Y}} |\, J(\mathbf{y}) - p_i(\mathbf{y}) \,| \, d\varphi(\mathbf{y}) \,\to\, 0 \quad \text{as } i \to \infty. \quad (9)$$

Moreover, define the functions $(\tilde{p}_i)$ as follows: $\tilde{p}_0 := p_0$, and

$$\mathbf{y} \mapsto \tilde{p}_i(\mathbf{y}) \,:=\, \max\,[\, \tilde{p}_{i-1}(\mathbf{y}), p_i(\mathbf{y}) \,], \quad i = 1, 2, \ldots$$

Then $\tilde{p}_i \to J(\cdot)$, φ-almost uniformly.

An optimal solution μ^* of $\mathbf{P}$ encodes *all* information on the optimal solutions $\mathbf{x}^*(\mathbf{y})$ of $\mathbf{P}_\mathbf{y}$. For instance, let $\mathbf{B}$ be a given Borel set of $\mathbb{R}^n$. Then from Theorem 1,

$$\mathrm{Prob}\,(\mathbf{x}^*(\mathbf{y})\in\mathbf{B}) \,=\, \mu^*(\mathbf{B}\times\mathbb{R}^p) \,=\, \varphi(g^{-1}(B)),$$

with g as in Theorem 1(b).

Moreover from Theorem 1(c), any optimal or nearly optimal solution of $\mathbf{P}^*$ provides us with some polynomial lower approximation of the optimal value function $\mathbf{y}\mapsto J(\mathbf{y})$ that converges to $J(\cdot)$ in the $L_1(\varphi)$ norm. Moreover, one may also obtain a piecewise polynomial approximation that converges to $J(\cdot)$, φ-almost uniformly.

In [6] the first author has defined a (J+M)-hierarchy of semidefinite relaxations $(\mathbf{Q}_i)$ to approximate as closely as desired the optimal value ρ. In particular, the dual of each semidefinite relaxation $\mathbf{Q}_i$ provides a polynomial $q_i\in\mathbb{R}[\mathbf{y}]$ bounded above by $J(\mathbf{y})$, and $\mathbf{y}\mapsto\tilde{q}_i(\mathbf{y}):=\max_{\ell=1,\dots i}q_\ell(\mathbf{y})$ converges φ-almost uniformly to the optimal value function J, as $i\to\infty$. This last property is the rationale behind the heuristic developed below.

III. A "JOINT+MARGINAL" APPROACH

Let $\mathbb{N}^n_i:=\{\alpha\in\mathbb{N}^n\,:\,|\alpha|\leq i\}$ with $|\alpha|=\sum_i\alpha_i$. With a sequence $\mathbf{z}=(z_\alpha)$ indexed in the canonical basis $(\mathbf{x}^\alpha)$ of $\mathbb{R}[\mathbf{x}]$, let $L_\mathbf{z}:\mathbb{R}[\mathbf{x}]\to\mathbb{R}$ be the linear mapping:

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

Moment matrix: The moment matrix $\mathbf{M}_i(\mathbf{z})$ associated with a sequence $\mathbf{z}=(z_\alpha)$, $\alpha\in\mathbb{N}^n_{2i}$, has its rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha)$, and with entries.

$$\mathbf{M}_i(\mathbf{z})(\alpha,\beta)\,=\,L_\mathbf{z}(\mathbf{x}^{\alpha+\beta})\,=\,z_{\alpha+\beta},\quad\forall\,\alpha,\beta\in\mathbb{N}^n_i.$$

Localizing matrix: Let q be the polynomial $\mathbf{x}\mapsto q(\mathbf{x}):=\sum_u q_u\mathbf{x}^u$. The localizing matrix $\mathbf{M}_i(q\,\mathbf{z})$ associated with $q\in\mathbb{R}[\mathbf{x}]$ and a sequence $\mathbf{z}=(z_\alpha)$, has its rows and columns indexed in the canonical basis $(\mathbf{x}^\alpha)$, and with entries.

$$\begin{aligned}\mathbf{M}_i(q\,\mathbf{z})(\alpha,\beta) &= L_\mathbf{z}(q(\mathbf{x})\mathbf{x}^{\alpha+\beta})\\ &= \sum_{u\in\mathbb{N}^n}q_u z_{\alpha+\beta+u},\quad\forall\,\alpha,\beta\in\mathbb{N}^n_i.\end{aligned}$$

A sequence $\mathbf{z}=(z_\alpha)\subset\mathbb{R}$ is said to have a *representing* finite Borel measure supported on $\mathbf{K}$ if there exists a finite Borel measure μ such that

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

A. A "joint+marginal" approach

With $\{f,(g_j)_{j=1}^m\}\subset\mathbb{R}[\mathbf{x}]$, let $\mathbf{K}\subset\mathbb{R}^n$ be the basic compact semi-algebraic set

$$\mathbf{K}:=\{\mathbf{x}\in\mathbb{R}^n\,:\,g_j(\mathbf{x})\geq 0,\,j=1,\dots,m\},\tag{10}$$

and consider the polynomial optimization problem (1).

Let $\mathbf{Y}_k\subset\mathbb{R}$ be some interval $[\underline{x}_k,\overline{x}_k]$, assumed to be contained in the orthogonal projection of $\mathbf{K}$ into the x_k-ccordinate axis.

For instance when the g_j's are affine (so that $\mathbf{K}$ is a convex polytope), $\underline{x}_k$ (resp. $\overline{x}_k$) solves the linear program $\min(\text{resp}\max)\,\{x_k\,:\,\mathbf{x}\in\mathbf{K}\}$. Similarly, when $\mathbf{K}$ is convex and defined by concave polynomials, one may obtain $\underline{x}_k$ and $\overline{x}_k$, up to (arbitrary) fixed precision. In many cases, (upper and lower) bound constraints on the variables are already part of the problem definition.

Let φ_k the probability measure uniformly distributed on $\mathbf{Y}_k$, hence with moments (β_ℓ) given by:

$$\beta_\ell=\int_{\underline{x}_1}^{\overline{x}_1}x^k d\varphi_k(x)\,=\,\frac{\overline{x}_k^{\ell+1}-\underline{x}_k^{\ell+1}}{(k+1)(\overline{x}_k-\underline{x}_k)}\tag{11}$$

for every $\ell=0,1,\dots$. Define the following parametric polynomial program in $n-1$ variables:

$$J^k(y)\,=\,\min_\mathbf{x}\{f(\mathbf{x})\,:\,\mathbf{x}\in\mathbf{K};\;x_k=y\},\tag{12}$$

or, equivalently $J^k(y)=\min\{f(\mathbf{x})\,:\,\mathbf{x}\in\mathbf{K}_y\}$, where for every $y\in\mathbf{Y}$:

$$\mathbf{K}_y:=\{\mathbf{x}\in\,\mathbf{K};\,x_k=y\}.\tag{13}$$

Observe that by definition, $f^*\,=\,\min_x\{J^k(x)\,:\,x\,\in\,\mathbf{Y}_k\}$, and $\mathbf{K}_y\neq\emptyset$ whenever $y\in\mathbf{Y}_k$, where $\mathbf{Y}_k$ is the orthogonal projection of $\mathbf{K}$ into the x_k-coordinate axis.

Semidefinite relaxations

To compute (or at least approximate) the optimal value ρ of problem $\mathbf{P}$ in (5) associated with the parametric optimization problem (12), we now provide a hierarchy of semidefinite relaxations in the spirit of those defined in [4]. Let $v_j:=\lceil(\deg \mathrm{g_j})/2\rceil$, $j=1,\dots,m$, and for $i\geq\max_j v_j$, consider the semidefinite program:

$$\begin{aligned}\rho_{ik}=\quad&\inf_\mathbf{z}\quad L_\mathbf{z}(f)\\ &\text{s.t.}\quad\mathbf{M}_i(\mathbf{z})\succeq 0,\,\mathbf{M}_{i-v_j}(g_j\,\mathbf{z})\succeq 0,\quad j=1,\dots,m\\ &\qquad L_\mathbf{z}(x_k^\ell)=\beta_\ell,\quad\ell=0,1,\dots 2i,\end{aligned}\tag{14}$$

where (β_ℓ) is defined in (11). We call (14) the *parametric semidefinite relaxation* of $\mathbf{P}$ with parameter $y=x_k$. Observe that without the "moment" constraints $L_\mathbf{z}(x_k^\ell)=\beta_\ell$, $\ell=1,\dots 2i$, the semidefinite program (14) is a relaxation of $\mathbf{P}$ and if $\mathbf{K}$ is compact, its corresponding optimal value f_i^* converges to f^* as $k\to\infty$; see Lasserre [4].

Letting $g_0\equiv 0$, the dual of (14) reads:

$$\begin{aligned}\rho^*_{ik}=\quad&\sup_{\lambda,(\sigma_j)}\sum_{\ell=0}^{2i}\lambda_\ell\,\beta_\ell\\ &\text{s.t.}\quad f(\mathbf{x})-\sum_{\ell=0}^{2i}\lambda_\ell x_k^\ell=\sigma_0+\sum_{j=1}^m\sigma_j\,g_j\\ &\qquad\sigma_j\in\Sigma[\mathbf{x}],\quad 0\leq j\leq m;\\ &\qquad\deg\sigma_j g_j\leq 2i,\quad 0\leq j\leq m.\end{aligned}\tag{15}$$

Equivalently, recall that $\mathbb{R}[x_k]_{2i}$ is the space of univariate polynomials of degree at most $2i$, and observe that in (15), the criterion reads

$$\sum_{\ell=0}^{2i}\lambda_\ell\,\beta_\ell=\int_{\mathbf{Y}_k}p_i(y)d\varphi_k(y),$$

where $p_i \in \mathbb{R}[x_k]_{2i}$ is the univariate polynomial $x_k \mapsto p_i(x_k) := \sum_{\ell=0}^{2i} \lambda_\ell x_k^\ell$. Then equivalently, the above dual may be rewritten as:

$$\begin{array}{rl} \rho_{ik}^* = & \displaystyle\sup_{p_i,(\sigma_j)} \int_{\mathbf{Y}_k} p_i d\varphi_k \\ \text{s.t.} & f - p_i = \sigma_0 + \displaystyle\sum_{j=1}^m \sigma_j\, g_j \\ & p_i \in \mathbb{R}[x_k]_{2i};\ \sigma_j \in \Sigma[\mathbf{x}], \quad 0 \leq j \leq m; \\ & \deg \sigma_j g_j \leq 2i, \quad 0 \leq j \leq m. \end{array} \tag{16}$$

Assumption 1: The family of polynomials $(g_j) \subset \mathbb{R}[\mathbf{x}]$ is such that for some $M > 0$,

$$\mathbf{x} \mapsto M - \|\mathbf{x}\|^2 = \sigma_0 + \sum_{j=1}^m \sigma_j\, g_j,$$

for some M and some s.o.s. polynomials $(\sigma_j) \subset \Sigma[\mathbf{x}]$.

Theorem 2: Let $\mathbf{K}$ be as (10) and Assumption 1 hold. Let the interval $\mathbf{Y}_k \subset \mathbb{R}$ be the orthognal projection of $\mathbf{K}$ into the x_k-coordinate axis, and let φ_k be the probability measure, uniformly distributed on $\mathbf{Y}_k$. Assume that $\mathbf{K}_y$ in (13) is not empty, let $y \mapsto J^k(y)$ be as in (12) and consider the semidefinite relaxations (14)-(16). Then as $i \to \infty$:

(a) $\rho_{ik} \uparrow \int_{\mathbf{Y}_k} J^k d\varphi_k$ and $\rho_{ik}^* \uparrow \int_{\mathbf{Y}_k} J^k d\varphi_k$

(b) Let $(p_i, (\sigma_j^i))$ be a nearly optimal solution of (16), e.g. such that $\int_{\mathbf{Y}_k} p_i d\varphi_k \geq \rho_{ik}^* - 1/i$. Then $p_i(y) \leq J^k(y)$ for all $y \in \mathbf{Y}_k$, and

$$\int_{\mathbf{Y}_k} |J^k(y) - p_i(y)|\, d\varphi_k(y) \;\to\; 0, \quad \text{as } i \to \infty. \tag{17}$$

Moreover, if one defines $\tilde{p}_0 := p_0$, and

$$\mathbf{y} \mapsto \tilde{p}_i(y) \;:=\; \max\,[\,\tilde{p}_{i-1}(y), p_i(y)\,], \quad i = 1, 2, \ldots,$$

then $\tilde{p}_i(y) \uparrow J^k(y)$, for φ_k-almost all $y \in \mathbf{Y}_k$, and so $\tilde{p}_i \to J^k$, φ_k-almost uniformly on $\mathbf{Y}_k$.

Theorem 2 is a direct consequence of [6, Corollary 2.6].

B. A "joint+marginal" algorithm for the general case

Theorem 2 provides a rationale for the following (J+M)-algorithm in the general case. In what follows we use the primal and dual semidefinite relaxations (14)-(15) with index i *fixed*.

ALGO 1: (J+M)-algorithm: non convex K, relaxation i

Set $k = 1$**;**
Step k: **Input:** $\mathbf{K}$, f, and the orthogonal projection $\mathbf{Y}_k = [\underline{x}_k, \overline{x}_k]$ of $\mathbf{K}$ into the x_k-coordinate axis, with associated probability measure φ_k, uniformly distributed on $\mathbf{Y}_k$.
Ouput: $\tilde{x}_k \in \mathbf{Y}_k$.

Solve the semidefinite program (16) and from an optimal (or nearly optimal) solution $(p_i, (\sigma_j))$ of (16), get a global minimizer $\tilde{x}_k$ of the univariate polynomial p_i on $\mathbf{Y}_k$.
If $k = n$ stop and output $\tilde{\mathbf{x}} = (\tilde{x}_1, \ldots, \tilde{x}_n)$, otherwise set $k = k + 1$ and repeat.

Of course, in general the vector $\tilde{\mathbf{x}} \in \mathbb{R}^n$ does not belong to $\mathbf{K}$. Therefore a final step consists of computing a local minimum $\hat{\mathbf{x}} \in \mathbf{K}$, by using some local minimization algorithm starting with the (unfeasible) initial point $\tilde{\mathbf{x}}$. Also note that when $\mathbf{K}$ is not convex, the determination of bounds $\underline{x}_k$ and $\overline{x}_k$ for the interval $\mathbf{Y}_k$ may not be easy, and so one might be forced to use a subinterval $\mathbf{Y}'_k \subseteq \mathbf{Y}_k$ with conservative (but computable) bounds $\underline{x}'_k \geq \underline{x}_k$ and $\overline{x}'_k \leq \overline{x}_k$.

Remark 1: Theorem 2 assumes that for every $y \in \mathbf{Y}_k$, the set $\mathbf{K}_y$ in (13) is not empty, which is the case if $\mathbf{K}$ is connected. If $\mathbf{K}_y = \emptyset$ for y in some open subset of $\mathbf{Y}_k$, then the semidefinite relaxation (14) has no solution ($\rho_{ik} = +\infty$), in which case one proceeds by dichotomy on the interval $\mathbf{Y}_k$ until $\rho_{ik} < \infty$.

C. A "joint+marginal" algorithm when $\mathbf{K}$ *is convex*

In this section, we now assume that the feasible set $\mathbf{K} \subset \mathbb{R}^n$ of problem $\mathbf{P}$ is convex (and compact). The idea is to compute $\tilde{x}_1$ as in **ALGO 1** and then repeat the procedure but now for the $(n-1)$-variable problem $\mathbf{P}(\tilde{x}_1)$ which is problem $\mathbf{P}$ in which the variable x_1 is *fixed* at the value $\tilde{x}_1$. This alternative is guaranteed to work if $\mathbf{K}$ is convex (but not always if $\mathbf{K}$ is not convex).

For every $j \geq 2$, denote by $\mathbf{x}_j \in \mathbb{R}^{n-j+1}$ the vector $(x_j, \ldots, x_n)$, and by $\tilde{\mathbf{x}}_{j-1} \in \mathbb{R}^{j-1}$ the vector $(\tilde{x}_1, \ldots, \tilde{x}_{j-1})$ (and so $\tilde{\mathbf{x}}_1 = \tilde{x}_1$).

Let the interval $\mathbf{Y}_1 \subset \mathbb{R}$ be the orthogonal projection of $\mathbf{K}$ into the x_1-coordinate axis. For every $\tilde{x}_1 \in \mathbf{Y}_1$, let the interval $\mathbf{Y}_2(\tilde{\mathbf{x}}_1) \subset \mathbb{R}$ be the orthogonal projection of the set $\mathbf{K} \cap \{\mathbf{x} : x_1 = \tilde{x}_1\}$ into the x_2-coordinate axis. Similarly, given $\tilde{\mathbf{x}}_2 \in \mathbf{Y}_1 \times \mathbf{Y}_2(\tilde{\mathbf{x}}_1)$, let the interval $\mathbf{Y}_3(\tilde{\mathbf{x}}_2) \subset \mathbb{R}$ be the orthogonal projection of the set $\mathbf{K} \cap \{\mathbf{x} : x_1 = \tilde{x}_1;\, x_2 = \tilde{x}_2\}$ into the x_3-coordinate axis, and etc. in the obvious way.

For every $k = 2, \ldots, n$, and $\tilde{\mathbf{x}}_{k-1} \in \mathbf{Y}_1 \times \mathbf{Y}_2(\tilde{\mathbf{x}}_1) \cdots \times \mathbf{Y}_{k-1}(\tilde{\mathbf{x}}_{k-2})$, let $\tilde{f}_k(\mathbf{x}_k) := f((\tilde{\mathbf{x}}_{k-1}, \mathbf{x}_k))$, and $\tilde{g}_j^k(\mathbf{x}_k) := g_j((\tilde{\mathbf{x}}_{k-1}, \mathbf{x}_k))$, $j = 1, \ldots, m$. Similarly, let

$$\begin{array}{rcl} \mathbf{K}_k(\tilde{\mathbf{x}}_{k-1}) & := & \{\mathbf{x}_k \,:\, \tilde{g}_j^k(\mathbf{x}_k) \geq 0,\ j = 1, \ldots, m\}, \\ & = & \{\mathbf{x}_k \,:\, (\tilde{\mathbf{x}}_{k-1}, \mathbf{x}_k) \in \mathbf{K}\}, \end{array} \tag{18}$$

and consider the problem:

$$\mathbf{P}(\tilde{\mathbf{x}}_{k-1}): \quad \min\,\{\tilde{f}_k(\mathbf{x}_x) \,:\, \mathbf{x}_x \in \mathbf{K}_j(\tilde{\mathbf{x}}_{k-1})\}, \tag{19}$$

i.e. the original problem $\mathbf{P}$ where the variable x_ℓ is fixed at the value $\tilde{x}_\ell$, for every $\ell = 1, \ldots, k-1$.

Write $\mathbf{Y}_j(\tilde{\mathbf{x}}_{k-1}) = [\underline{x}_k, \overline{x}_k]$, and let φ_k be the probability measure uniformly distributed on $\mathbf{Y}_k(\tilde{\mathbf{x}}_{k-1})$.

Let $\mathbf{z}$ be a sequence indexed in the monomial basis of $\mathbb{R}[\mathbf{x}_k]$. With index i, fixed, the parametric semidefinite relaxation (14) with parameter x_k, associated with problem $\mathbf{P}(\tilde{\mathbf{x}}_{k-1})$, reads:

$$\begin{array}{rl} \rho_{ik} = \displaystyle\inf_{\mathbf{z}} & L_{\mathbf{z}}(\tilde{f}_k) \\ \text{s.t.} & \mathbf{M}_i(\mathbf{z}),\ \mathbf{M}_{i-v_j}(\tilde{g}_j^k\,\mathbf{z}) \succeq 0, \quad j = 1, \ldots, m \\ & L_{\mathbf{z}}(x_k^\ell) = \beta_\ell, \quad \ell = 0, 1, \ldots, 2i, \end{array} \tag{20}$$

where (β_ℓ) is defined in (11). Its dual is the semidefinite program (with $\tilde{g}_0^k \equiv 1$)):

$$\begin{aligned} \rho_{ik}^* \quad = \quad & \sup_{p_i,(\sigma_j)} \int_{\mathbf{Y}_k(\tilde{\mathbf{x}}_{k-1})} p_i d\varphi_k \qquad (21) \\ \text{s.t.} \quad & \tilde{f}_k - p_i = \sigma_0 + \sum_{j=1}^m \sigma_j \, \tilde{g}_j^k \\ & p_i \in \mathbb{R}[x_k]_{2i}, \; \sigma_j \in \Sigma[\mathbf{x}_k], \quad j = 0, \ldots, m \\ & \deg \sigma_j \tilde{g}_j^k \leq 2i, \quad j = 0, \ldots, m. \end{aligned}$$

The important difference between (14) and (20) is the *size* of the corresponding semidefinite programs, since $\mathbf{z}$ in (14) (resp. in (20)) is indexed in the canonical basis of $\mathbb{R}[\mathbf{x}]$ (resp. $\mathbb{R}[\mathbf{x}_k]$).

The (J+M)-algorithm for $\mathbf{K}$ *convex*

Recall that the order i of the semidefinite relaxation is fxed. The (J+M)-algorithm consists of n steps. At step k of the algorithm, the vector $\tilde{\mathbf{x}}_{k-1} = (\tilde{x}_1, \ldots, \tilde{x}_{k-1})$ (already computed) is such that $\tilde{x}_1 \in \mathbf{Y}_1$ and $\tilde{x}_\ell \in \mathbf{Y}_\ell(\tilde{\mathbf{x}}_{\ell-1})$ for every $\ell = 2, \ldots, k-1$, and so the set $\mathbf{K}_k(\tilde{\mathbf{x}}_{k-1})$ is a nonempty compact convex set.

ALGO 2: (J+M)-algorithm: convex K, relaxation i

Set $k = 1$**;**
Step $k \geq 1$: **Input:** For $k = 1$, $\tilde{\mathbf{x}}_0 = \emptyset$, $\mathbf{Y}_1(\tilde{\mathbf{x}}_0) = \mathbf{Y}_1$; $\mathbf{P}(\tilde{\mathbf{x}}_0) = \mathbf{P}$, $f_1 = f$ and $\tilde{g}_j^1 = g_j$, $j = 1, \ldots, m$.
For $k \geq 2$, $\tilde{\mathbf{x}}_{k-1} \in \mathbf{Y}_1 \times \mathbf{Y}_2(\tilde{x}_1) \cdots \times \mathbf{Y}_{k-1}(\tilde{x}_{k-2})$.
Output: $\tilde{\mathbf{x}}_k = (\tilde{\mathbf{x}}_{k-1}, \tilde{x}_k)$ with $\tilde{x}_k \in \mathbf{Y}_k(\tilde{\mathbf{x}}_{k-1})$.
Consider the parametric semidefinite relaxations (20)-(21) with parameter x_k, associated with problem $\mathbf{P}(\tilde{\mathbf{x}}_{k-1})$ in (19).

- From an optimal solution of (21), extract the univariate polynomial $x_k \mapsto p_i(x_k) := \sum_{\ell=0}^{2i} \lambda_\ell^* x_k^\ell$.
- Get a global minimizer $\tilde{x}_k$ of p_i on the interval $\mathbf{Y}_k(\tilde{\mathbf{x}}_{k-1}) = [\underline{x}_k, \overline{x}_k]$, and set $\tilde{\mathbf{x}}_k := (\tilde{\mathbf{x}}_{k-1}, \tilde{x}_k)$.

If $k = n$ stop and ouput $\tilde{\mathbf{x}} \in \mathbf{K}$, otherwise set $k = k + 1$ and repeat.

As $\mathbf{K}$ is convex, $\tilde{\mathbf{x}} \in \mathbf{K}$ and one may stop. A refinement is to now use $\tilde{\mathbf{x}}$ as the initial guess of a local minimization algorithm to obtain a local minimizer $\hat{\mathbf{x}} \in \mathbf{K}$ of $\mathbf{P}$. In view of Theorem 2, the larger the index i of the relaxations (20)-(21), the better the values $f(\tilde{\mathbf{x}})$ and $f(\hat{\mathbf{x}})$.

Of course, **ALGO 2** can also be used when $\mathbf{K}$ is not convex. However, it may happen that at some stage k, the semidefinite relaxation (20) may be infeasible because $J^k(y)$ is infinite for some values of $y \in \mathbf{Y}_k(\tilde{\mathbf{x}}_{k-1})$. This is because the feasible set $\mathbf{K}(\tilde{\mathbf{x}}_{k-1})$ in (18) may be disconnected.

IV. Computational experiments

We report on preliminary computational experiments on some non convex NP-hard optimization problems. We have tested the algorithms on a set of difficult global optimization problems taken from Floudas et al. [1]. To solve the semidefinite programs involved in **ALGO 1** and in **ALGO 2**, we have used the GloptiPoly software [3] that implements the hierarchy of semidefinite relaxations defined in [4, (4.5)].

Prob	n	m	f^*	i	ALGO 2	rel. error
2.2	5	11	-17	2	-17.00	0%
2.3	6	8	-361.5	1	-361.50	0%
2.6	10	21	-268.01	1	-267.00	0.3%
2.9	10	21	0	1	0.00	0%
2.8C1	20	30	-394.75	1	-385.30	2.4%
2.8C2	20	30	-884.75	1	-871.52	1.5%
2.8C3	20	30	-8695	1	-8681.7	0.15%
2.8C4	20	30	-754.75	1	-754.08	0.09%
2.8C5	20	30	-4150.41	1	-3678.2	11%

TABLE I
ALGO 2 FOR CONVEX SET $\mathbf{K}$

A. ALGO 2 *for convex set* $\mathbf{K}$

Those problems are taken from [1, §2]. The set $\mathbf{K}$ is a convex polytope and the function f is a nonconvex quadratic polynomial $\mathbf{x} \mapsto \mathbf{x}'Q\mathbf{x} + \mathbf{b}'\mathbf{x}$ for some real symmetric matrix Q and vector $\mathbf{b}$. In Table I one displays the problem name, the number n of variables, the number m of constraints, the gobal optimum f^*, the index i of the semidefinite relaxation in **ALGO 2**, the optimal value obtained using the output of **ALGO 2** as initial guess in a local minimization algorithm of the MATLAB toolbox, and the associated relative error. As recommended in Gloptipoly [3] for numerical stability and precision, the problem data have been rescaled to obtain a polytope contained in the box $[-1, 1]^n$. As one may see, and excepted for problem 2.8C5, the relative error is very small. For the last problem the relative error (about 11%) is relatively high despite enforcing some extra upper and lower bounds $\underline{x_i} \leq x_i \leq \overline{x}_i$, after reading the optimal solution. However, using $\tilde{\mathbf{x}} \in \mathbf{K}$ as initial guess of the local minimization algorithm in MATLAB, one still finds the optimal value f^*.

B. ALGO 1 *for non convex set* $\mathbf{K}$

In this section, $\mathbf{K} = \{\mathbf{x} : \mathbf{x} \in \Omega; \, \mathbf{x}'Q\mathbf{x} + \mathbf{b}'\mathbf{x} \geq 0\}$ where Ω is a convex polytope and Q is neither semidefinite nor negative definite. Again in tables below, n (resp. m) stands for the number of variables (resp. constraints), and the value displayed in the "**ALGO 1**" column is obtained in running a local minimization algorithm of the MATLAB toolbox with the output $\tilde{\mathbf{x}}$ of **ALGO 1** as initial guess.

In Problems 3.2, 3.3 and 3.4 from Floudas et al. [1, §3], one has $2n$ linear bound constraints and additional linear and non convex quadratic constraints. As one may see, the results displayed in Table II are very good.

For the Haverly Pooling problem 5.2.2 in [1, §5] with three different data sets, one has $n = 9$ and $m = 24$ constraints, among which 3 nonconvex bilinear constraints and 18 linear bound constraints $0 \leq x_i \leq 500$, $i = 1, \ldots, 9$. In the first run of **ALGO 1** we obtained bad results because the bounds are very loose and in the hierarchy of lower bounds (f_k^*) in [4] that converge to f^*, if on the one hand $f_2^* = f^*$, on the other hand the lower bound $f_1^* < f^*$ is loose. In such a case, and in view of the rationale behind the "joint+marginal" approach, it is illusory to obtain good results with **ALGO**

Prob	n	m	f^*	i	ALGO 1	rel. error
3.2	8	22	7049	1	7049	0%
3.3	5	16	-30665	1	-30665	0%
3.4	6	18	-310	1	-298	3.8%

TABLE II
ALGO 1 FOR NON CONVEX SET $\mathbf{K}$

Prob	n	m	f^*	i	ALGO 1	rel. error
5.2.2 (1)	9	24	400	1	400	0%
5.2.2 (2)	9	24	600	1	600	0%
5.2.3 (3)	9	24	750	1	750	0%

TABLE III
ALGO 1 FOR NON CONVEX SET $\mathbf{K}$

1 or **ALGO 2**. Therefore, from the optimal solution $\mathbf{x}^*$ in [1], and when $0 < x_i^* < 500$, we have generated stronger bounds $0.4x_i^* \leq x_i \leq 1.6x_i^*$. In this case, f_1^* is much closer to f^* and we obtain the global minimum f^* with **ALGO 1** followed by the local minimization subroutine; see Table III. Importantly, in **ALGO 1**, and before running the local optimization subroutine, one ends up with a non feasible point $\tilde{\mathbf{x}}$. Moreover, we had to sometimes use the dichotomy procedure of Remark 1 because if $\mathbf{Y}_k$ is large, one may have $\mathbf{K}_y = \emptyset$ for y in some open subintervals of $\mathbf{Y}_k$.

Finally, in problem 5.2.4 in [1] we also obtained the global optimum f^*. In Problem 7.2.2, to handle the non-polynomial function $x_i^{0.5}$, one uses the lifting $u_i^2 = x_i$, $u_i \geq 0$, $i = 5, 6$. Here again, one obtains the optimal value f^* with **ALGO 1** followed by a local optimization subroutine.

C. **ALGO 2** *for MAXCUT*

Finally we have tested **ALGO 2** on the famous NP-hard discrete optimization problem MAXCUT, which consists of minimizing a quadratic form $\mathbf{x} \mapsto \mathbf{x}'Q\mathbf{x}$ on $\{-1, 1\}^n$, for some real symmetric matrix $Q \in \mathbb{R}^{n\times n}$. In this case, $\mathbf{Y}_k = \{-1, 1\}$ and the marginal constraint $L_{\mathbf{z}}(x_k^\ell) = \gamma_\ell$ in (20) need only be imposed for $\ell = 1$, because of the constraints $x_k^2 = 1$ for every $k = 1, \ldots, n$. Accordingly, in an optimal solution of the dual (21), $p_i \in \mathbb{R}[x_k]$ is an affine polynomial $x_k \mapsto p_i(x_k) = \lambda_0 + \lambda_1 x_k$ for some scalars λ_0, λ_1. Therefore after solving (21) one decides $\tilde{x}_k = -1$ if $p_i(-1) < p_i(1)$ (i.e. if $\lambda_1 > 0$) and $\tilde{x}_k = 1$ otherwise.

Recall that in **ALGO 2** one first compute $\tilde{x}_1$, then with x_1 fixed at the value $\tilde{x}_1$, one computes $\tilde{x}_2$, etc. until one finally computes $\tilde{x}_n$, and get $\tilde{\mathbf{x}}$. In what we call the "max-gap" variant of **ALGO 2**, one first solves n programs (14)-(15) with parameter x_1 to obtain an optimal solution $p_i(x_1) = \lambda_0^1 + \lambda_1^1 x_1$ of the dual (15), then with x_2 to obtain $(\lambda_0^2, \lambda_1^2)$, etc. finally with x_n to obtain $(\lambda_0^n, \lambda_1^n)$. One then select k such that $|\lambda_1^k| = \max_\ell |\lambda_1^\ell|$, and compute $\tilde{x}_k$ accordingly. This is because the larger $|\lambda_1|$, (i.e. the larger $|p_i(-1) - p_i(1)|$), the more likely the choice -1 or 1 is correct. After x_k is fixed at the value $\tilde{x}_k$, one repeats the procedure for the $(n-1)$-problem $\mathbf{P}(\tilde{x}_k)$, etc.

We have tested the "max-gap" variant for MAXCUT problems on random graphs with $n = 20, 30$ and 40 nodes. For each value of n, we have solved 50 randomly generated problems and 100 for $n = 40$. The probability φ_k on $\mathbf{Y}_k = \{-1, 1\}$ is uniform (i.e., $\beta_1 = 0$ in (20)). Let f_1^* denote the optimal value of the Shor's relaxation with famous Goemans and Williamson's 0.878 performance guarantee. Let ρ denote the cost of the solution $\mathbf{x} \in \{-1, 1\}^n$ generated by the **ALGO 2**. In Table IV below, we have reported the average relative error $(\rho - f_1^*)/|f_1^*|$, which as one may see, is comparable with the Goemans and Williamson (GW) ratio.

n	20	30	40
$(\rho - f_1^*)/\lvert f_1^*\rvert$	10.3%	12.3%	12.5%

TABLE IV
RELATIVE ERROR FOR MAXCUT

V. CONCLUSION

First preliminary results are promising, even with small relaxation order i. When the feasible set is non convex, it may become difficult to obtain a feasible solution and an interesting issue for further investigation is how to proceed when $\mathbf{K}_y = \emptyset$ for y in some open subinterval of $\mathbf{Y}_k$ (proceeding by dichotomy on $\mathbf{Y}_k$ is one possiblity).

REFERENCES

[1] C.A. Floudas et al., *Handbook of Test Problems in Local and Global optimization*, Kluwer Academic Publishers, Dordrecht, 1999.

[2] M.X. Goemans, D.P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, Journal of the ACM 42, pp.1115-1145, 1995.

[3] D. Henrion, J. B. Lasserre, J. Lofberg, *GloptiPoly 3: moments, optimization and semidefinite programming*, Optim. Methods and Softw. 24, pp. 761–779, 2009. `http://www.laas.fr/~henrion/software/gloptipoly3/`

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

[5] J.B. Lasserre, *Polynomial programming: LP-relaxations also converge*, SIAM J. Optim. 15, pp. 383–393, 2004.

[6] J.B. Lasserre, *A "joint+marginal" approach to parametric polynomial optimization*, SIAM J. Optim. 20, pp. 1995-2022, 2010.