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 $P$ 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 $P(x_1)$ where the parameter $x_1$ is fixed and takes values in some interval $Y_1 \subseteq \mathbb{R}$, with some probability $\varphi_1$ uniformly distributed on $Y_1$. 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 $P(x_1)$, as $k$ increases. Then with $k$ fixed a priori, one computes $\hat{x}_1 \in Y_1$ which minimizes the univariate polynomial $p_k(x_1)$ on the interval $Y_1$, 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 $P(x_2)$ with parameter $x_2$ in some new interval $Y_2 \subseteq \mathbb{R}$, etc. so as to finally obtain a vector $\tilde{x} \in \mathbb{R}^n$. Preliminary numerical results are provided.

I. INTRODUCTION

Consider the general polynomial program

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

where $f$ is a polynomial, $K \subseteq \mathbb{R}^n$ is a basic semi-algebraic set, and $f^*$ is the global minimum of $P$ (as opposed to a local minimum). One way to approximate the global optimum $f^*$ of $P$ 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 reveal 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 results 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 $x \in \mathbb{R}^n$ and parameters $y$ in a simple set $Y$, consists of the standard hierarchy of semidefinite relaxations in [4] where one treats the parameters $y$ also as variables. But now the moment-approach implemented in the semidefinite relaxations, considers a joint probability distribution on the pair $(x, y)$, with the additional constraint that the marginal distribution on $Y$ is fixed (e.g. the uniform probability distribution on $Y$); whence the name "joint+marginal".

For every $k = 1, \ldots, n$, let the compact interval $Y_k := [\bar{y}_k, \bar{y}_k] \subseteq \mathbb{R}$ be contained in the projection of $K$ into the $x_k$-coordinate axis. In the context of the (non-parametric) polynomial optimization ([6]), 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 $Y_1 = [\bar{x}_1, \bar{x}_1]$ with associated probability distribution $\varphi_1$ uniformly distributed on $Y_1$.
- (b) with $i \in \mathbb{N}$ fixed, solve the $i$-th semidefinite relaxation of the (J+M)-hierarchy [6] applied to problem $P(x_1)$ with $n-1$ variables $(x_2, \ldots, x_n)$ and parameter $x_1$, which is problem $P$ with the additional constraint that the variable $x_1 \in Y_1$ is fixed. The dual provides a univariate polynomial $x_1 \mapsto J^i_1(x_1)$ which, if $i$ would increase, would converge to $J^1_1(x_1)$ in the $L_1(\varphi_1)$-norm. (The map $v \mapsto J^1(v)$ denotes the optimal value function of $P(v)$, i.e. the optimal value of $P$ given that the variable $x_1$ is fixed at the value $v$.) Next, compute $\tilde{x}_1 \in Y_1$, a global minimizer of the univariate polynomial $J^i_1$ on $Y_1$ (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 $x^* = (x^*_1, \ldots, x^*_n)$ of $P$.
- (c) go back to step (b) with now $x_2 \in Y_2 \subseteq \mathbb{R}$ instead of $x_1$, and with $\varphi_2$ being the probability measure uniformly distributed on $Y_2$. With the same method, compute a global minimizer $\tilde{x}_2 \in Y_2$, of the univariate polynomial $x_2 \mapsto J^2_2(x_2)$ on the interval $Y_2$. 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 $P(x_2)$ (i.e. the optimal value of $P$ given that the variable $x_2$ is fixed at the value $v$). Iterate until one has obtained $\tilde{x}_n \in Y_n \subseteq \mathbb{R}$.

One ends up with a point $\tilde{x} \in \prod_{k=1}^{n} Y_k$ and in general $\tilde{x} \not\in K$. One may then use $\tilde{x}$ as initial guess of a local optimization procedure to find a local minimum $\hat{x} \in K$. 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

This work is supported by the ANR project "Semidefinite Hierarchy for Polynomial Optimization with Application to Sum-of-Squares" (Oxygene).
The rational behind the (J+M)-algorithm is that if \( i \) is large enough and \( P \) has a unique global minimizer \( x^* \in K \), then \( \tilde{x} \) as well as \( \tilde{x} \) should be close to \( x^* \).

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 the same order as the \( i \)-th semidefinite relaxation in the hierarchy defined in [4]), i.e., a polynomial in the input size of \( P \).

When the feasible set \( K \) is convex, one may define the following variant to obtain a feasible point \( \tilde{x} \in K \). Again, let \( Y_1 \) be the projection of \( K_1 \) into the \( x_1 \)-coordinate axis. Once \( \tilde{x}_1 \in Y_1 \) is obtained in step (b), consider the new optimization problem \( P(\tilde{x}_1) \) in the \( n-1 \) variables \( (x_2, \ldots, x_n) \), obtained from \( P \) by fixing the variable \( x_1 \in Y_1 \) at the value \( \tilde{x}_1 \). Its feasible set is the convex set \( K_1 := K \cap \{ x : x_1 = \tilde{x}_1 \} \). Let \( Y_2 \) be the projection of \( K_1 \) into the \( x_2 \)-coordinate axis. Then go back to step (b) with now \( x_2 \in Y_2 \) as parameter and \( (x_1, \ldots, x_n) \) as variables, to obtain a point \( \tilde{x}_2 \in Y_2 \), etc. until a point \( \tilde{x} \in \prod_{i=1}^{n} Y_i \) is obtained. Notice that now \( \tilde{x} \in K \) because \( K \) is convex. Then proceed as before with \( \tilde{x} \) being the initial guess of a local minimization algorithm to obtain a local minimizer \( \tilde{x} \in K \) of \( P \).

II. THE ”JOINT+MARGINAL APPROACH TO PARAMETRIC
OPTIMIZATION

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

The parametric optimization problem

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

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

and for each \( y \in Y \), let

\[
K_y := \{ x \in \mathbb{R}^n : (x, y) \in K \}. \quad (3)
\]

For each \( y \in Y \), fixed, consider the optimization problem:

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

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

For instance

\[
\varphi(B) := \left( \int_Y dy \right)^{-1} \int_{Y \cap B} dy , \quad \forall B \in \mathcal{B}(\mathbb{R}^p),
\]

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

A. A related infinite-dimensional linear program

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

\[
\rho := \inf_{\mu \in M(K)} \left\{ \int_K f \, d\mu : \mu = \varphi \right\} , \quad (5)
\]

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

The dual of \( P \) is the following infinite-dimensional linear program:

\[
\rho^* := \sup_{p \in P[Y]} \int_Y p(y) d\varphi(y) \quad \text{subject to } f(x) - p(y) \geq 0 \quad \forall (x, y) \in K . \quad (6)
\]

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

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

(a) \( \rho = \int_Y J(y) d\varphi(y) \) and \( P \) has an optimal solution.

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

\[
g(y) = x^*(y) \quad \text{for every } y \in Y .
\]

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

\[
\int_K x^\alpha y^\beta d\mu(x,y) = \int_Y y^\beta g(y)^\alpha d\varphi(y) . \quad (8)
\]

(c) There is no duality gap between (5) and (6), i.e. \( \rho = \rho^* \), and if \( (p_i) \in \mathbb{N}^Y \) is a maximizing sequence of (6) then:

\[
\int_Y | J(y) - p_i(y) | d\varphi(y) \to 0 \quad \text{as } i \to \infty . \quad (9)
\]

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

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

Then \( \tilde{p}_{\infty} \to J(\cdot) \), \( \varphi \)-almost uniformly.
An optimal solution $\mu^* \in P$ encodes all information on the optimal solutions $x^*(y)$ of $P_y$. For instance, let $B$ be a given Borel set of $\mathbb{R}^n$. Then from Theorem 8

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

with $g$ as in Theorem 8(b).

Moreover from Theorem 8(c), any optimal or nearly optimal solution of $P^*$ provides us with some polynomial lower approximation of the optimal value function $y \mapsto J(y)$ that converges to $J(\cdot)$ in the $L_1(\varphi)$ norm. Moreover, one may also obtain a piecewise polynomial approximation that converges to $J(\cdot)$, $\varphi$-almost uniformly.

In [6] the first author has defined a (J+M)-hierarchy of semidefinite relaxations $Q_\ell$ to approximate as closely as desired the optimal value $\rho$. In particular, the dual of each semidefinite relaxation $Q_\ell$ provides a polynomial $q_\ell \in \mathcal{R}[y]$ bounded above by $J(y)$, and $y \mapsto q_\ell(y) := \max_{i=1,...,I} q_i(y)$ converges $\varphi$-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 $\mathbb{N}^n := \{\alpha \in \mathbb{N}^n : \|\alpha\| \leq i\}$ with $\|\alpha\| := \sum_i \alpha_i$. With a sequence $z = (z_\alpha)$ indexed in the canonical basis ($x^\alpha$) of $\mathbb{R}[x]$, let $L_z : \mathbb{R}[x] \to \mathbb{R}$ be the linear mapping:

$$f(\sum_\alpha f_\alpha x^\alpha) \mapsto L_z(f) := \sum_\alpha f_\alpha z_\alpha, \quad f \in \mathbb{R}[x].$$

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

$$M_z(z)(\alpha, \beta) := L_z(x^{\alpha+\beta}) = z_{\alpha+\beta}, \quad \forall \alpha, \beta \in \mathbb{N}^n_{2i}.$$

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

$$M_z(q : z)(\alpha, \beta) = L_z(q(x) x^{\alpha+\beta}) = \sum_u q_u z_{\alpha+\beta+u}, \quad \forall \alpha, \beta \in \mathbb{N}^n.$$

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

$$z_\alpha = \int_K x^\alpha \mu dx, \quad \forall \alpha \in \mathbb{N}^n_{2i}.$$

A "joint+ marginal" approach

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

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

and consider the polynomial optimization problem [1].

Let $Y_k \subset \mathbb{R}$ be some interval $[x_k, \bar{x}_k]$, assumed to be contained in the orthogonal projection of $K$ into the $x_k$-coordinate axis.

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

Let $\varphi_k$ the probability measure uniformly distributed on $Y_k$, hence with moments ($\beta_k$) given by:

$$\beta_k = \int_{x_k}^y x^\ell d\varphi_k(x) = \frac{x_k^{\ell+1} - \bar{x}_k^{\ell+1}}{(k+1)(\bar{x}_k - x_k)}$$

for every $\ell = 0, 1, \ldots, \text{define the following parametric polynomial program in } n - 1 \text{ variables:}$

$$J^k(y) = \min \{f(x) : x \in K ; \ x_k = y\},$$

or, equivalently $J^k(y) = \min \{f(x) : x \in K_{y}\}$, where for every $y \in Y_k$:

$$K_y := \{x \in K ; x_k = y\}.$$

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

**Semidefinite relaxations**

To compute (or at least approximate) the optimal value $\rho$ of problem $P$ in [8] associated with the parametric optimization problem [12], we now provide a hierarchy of semidefinite relaxations in the spirit of those defined in [4].

Let $v_j := [(\text{deg } g_j)/2], j = 1, \ldots, m$, and for $i \geq \max_j v_j$, consider the semidefinite program:

$$\rho_{ik} = \inf_z L_x(f)$$

s.t. $M_z(z) \succeq 0, M_{x_i - v_j}(g_j z) \succeq 0, \ j = 1, \ldots, m$

$$L_x(x^\ell_k) = \beta_k, \ell = 0, 1, \ldots, 2i,$$

where ($\beta_k$) is defined in [11]. We call [14] the parametric semidefinite relaxation of $P$ with parameter $y = x_k$. Observe that without the "moment" constraints $L_x(x^\ell_k) = \beta_k, \ell = 1, \ldots, 2i$, the semidefinite program [14] is a relaxation of $P$ and if $K$ is compact, its corresponding optimal value $f^*$ converges to $f^*$ as $k \to \infty$: see Lasserre [4].

Letting $g_0 \equiv 0$, the dual of [14] reads:

$$\rho^*_{ik} = \sup_{\lambda \in \mathbb{R}[x]} \sum_{\ell=0}^{2i} \lambda_\ell \beta_k$$

s.t. $f(x) - \sum_{\ell=0}^{2i} \lambda_\ell x^\ell_k = \sigma_0 + \sum_{j=1}^m \sigma_j g_j$

$$\sigma_j \in \mathbb{R}[x], \ 0 \leq j \leq m;$$

$$\sigma_0, \sigma_j g_j \leq 2i, \ 0 \leq j \leq m.$$
where \( p_i \in \mathbb{R}[x_k]_{≤ 2} \) is the univariate polynomial \( x_k \mapsto p_i(x_k) := \sum_{j=0}^{m} \lambda_j x_k^j \). Then equivalently, the above dual may be rewritten as:

\[
\rho_{ik} = \sup_{p_i, (\sigma_j)} \int_{Y_k} p_i d\varphi_k \quad \text{s.t.} \quad f - p_i = \sigma_0 + \sum_{j=1}^{m} \sigma_j g_j \tag{16}
\]

for some \( \sigma_0 \) and \( \sigma_j \in \mathbb{R} \). Let \( \rho_{ik} \) be as \( \rho_{ik} \) and Assumption 1 hold. The family of polynomials \( \{g_j\} \subset R^n \) is not convex, the determination of bounds \( y_{ik}, z_{ik} \) for the interval \( Y_k \) may not be easy, and so one might be forced to use a subinterval \( Y_{ik} \subset Y_k \) with conservative (but computable) bounds \( y_{ik}, \rho_{ik} \) until \( \rho_{ik} < \infty \).

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

**C. A "joint-marginal" algorithm when K is convex**

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

For every \( j \geq 2 \), denote by \( x_{j-1} \in \mathbb{R}^{n−j+1} \) the vector \((x_j, \ldots, x_n)\), and by \( x_{j-1} \in \mathbb{R}^{n−j+1} \) the vector \((x_1, \ldots, x_{j-1})\) (and so \( x_1 = \bar{x}_1 \)).

Let the interval \( Y_1 \subset \mathbb{R} \) be the orthogonal projection of \( K \) into the \( x_1 \)-coordinate axis. For every \( \bar{x}_1 \in Y_1 \), let the interval \( Y_2(\bar{x}_1) \subset \mathbb{R} \) be the orthogonal projection of the set \( K \cap \{x : x_1 = \bar{x}_1\} \) into the \( x_2 \)-coordinate axis. Similarly, given \( x_2 \in Y_1 \times Y_2(\bar{x}_1) \), let the interval \( Y_3(x_2) \subset \mathbb{R} \) be the orthogonal projection of the set \( K \cap \{x : x_1 = \bar{x}_1; x_2 = \bar{x}_2\} \) into the \( x_3 \)-coordinate axis, and etc. in the obvious way.

For every \( k = 2, \ldots, n \), and \( \bar{x}_{k-1} \in Y_1 \times Y_2(\bar{x}_2) \) \( \cdots Y_{k-1}(\bar{x}_{k-2}) \), let \( \bar{f}_k(x_k) := f(\bar{x}_{k-1}, x_k) \), and \( \bar{g}^j_k(x_k) := g_j(\bar{x}_{k-1}, x_k), j = 1, \ldots, m \). Similarly, let

\[
K_k(\bar{x}_{k-1}) := \{ x_k : \bar{g}^j_k(x_k) ≥ 0, j = 1, \ldots, m \}, \tag{18}
\]

and consider the problem:

\[
P(\bar{x}_{k-1}) := \min \{ \bar{f}_k(x_k) : x_k \in K_k(\bar{x}_{k-1}) \}, \tag{19}
\]

i.e. the original problem \( P \) where the variable \( x_k \) is fixed at the value \( \bar{x}_k \), for every \( \bar{x} = (\bar{x}_1, \ldots, \bar{x}_k) \).

Write \( Y_j(\bar{x}_{k-1}) = (\bar{x}_j, \bar{\tau}_j) \), and let \( \varphi_k \) be the probability measure uniformly distributed on \( Y_k(\bar{x}_{k-1}) \).

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

\[
\rho_{ik} = \inf_{\rho} L_k(\bar{f}_k) \quad \text{s.t.} \quad \rho_{M}^{(j)}(z) \geq 0, \quad j = 1, \ldots, m
\]

\[
\rho_{ik} = \inf_{\rho} L_k(\bar{f}_k) \quad \text{s.t.} \quad \rho_{M}^{(j)}(z) \geq 0, \quad j = 1, \ldots, m
\]

\[
L_k(\bar{x}_k) = \rho_{ik}, \quad \ell = 0, 1, \ldots, 2i, \tag{20}
\]
where \((\beta_t)\) is defined in \((11)\). Its dual is the semidefinite program (with \(\tilde{g}_0 \equiv 1\)):

\[
\rho_{jk}^p = \sup_{p_i(\sigma_j)} \int Y_k(x_{k-1}) \, p_i \, d\varphi_k \tag{21}
\]

s.t. \(f_k - p_i = \sigma_0 + \sum_{j=1}^m \sigma_j \tilde{g}_j^k\)

\[
p_i \in \mathbb{R}[x_1, x_2, \ldots, x_m], \quad \sigma_j \in \mathbb{S}[x_k], \quad j = 0, \ldots, m
\]

\[
deg \sigma_j \tilde{g}_j^k \leq 2i, \quad j = 0, \ldots, m
\]

The important difference between \((14)\) and \((21)\) is the size of the corresponding semidefinite programs, since \(z \in \mathbb{R}[x]\) (resp. \(\mathbb{R}[x_k]\)) is indexed in the canonical basis of \(\mathbb{R}[x]\) (resp. \(\mathbb{R}[x_k]\)).

The \((J+M)\)-algorithm for \(K\) convex

Recall that the order \(i\) of the semidefinite relaxation is fixed. The \((J+M)\)-algorithm consists of \(n\) steps. At step \(k\) of the algorithm, the vector \(\tilde{x}_{k-1} = (\tilde{x}_1, \ldots, \tilde{x}_{k-1})\) (already computed) is such that \(\tilde{x}_1 \in Y_1\) and \(\tilde{x}_\ell \in Y_\ell(\tilde{x}_{\ell-1})\) for every \(\ell = 2, \ldots, k - 1\), and so the set \(K_k(\tilde{x}_{k-1})\) is a nonempty compact convex set.

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

Set \(k = 1; \)

\(\text{Step } k \geq 1: \text{Input: For } k = 1, \tilde{x}_0 = \emptyset, Y_1(\tilde{x}_0) = Y_1; \)

\(P(\tilde{x}_0) = P, f_1 = f \text{ and } g_j^1 = g_j, j = 1, \ldots, m.\)

For \(k \geq 2, \tilde{x}_{k-1} \in Y_1 \times Y_2(\tilde{x}_1) \times \cdots \times Y_{k-1}(\tilde{x}_{k-2}).\)

\(\text{Output: } \tilde{x}_k = (\tilde{x}_{k-1}, \tilde{x}_k) \text{ with } \tilde{x}_k \in Y_k(\tilde{x}_{k-1}).\)

Consider the parametric semidefinite relaxations \((20)-(21)\) with parameter \(x_k\), associated with problem \(P(\tilde{x}_{k-1})\) in \((19).\)

- From an optimal solution of \((21)\), extract the univariate polynomial \(x_k \to p_i(x_k) := \sum_{\ell=0}^n \lambda^*_\ell x_k^\ell.\)
- Get a global minimizer \(\tilde{x}_k\) of \(p_i\) on the interval \(Y_k(\tilde{x}_{k-1}) = [\underline{\tau}_k, \overline{\tau}_k]\), and set \(\tilde{x}_k := (\tilde{x}_{k-1}, \tilde{x}_k).\)

If \(k = n\) stop and output \(\tilde{x} \in K\), otherwise set \(k = k + 1\) and repeat.

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

Of course, ALGO 2 can also be used when \(K\) is not convex. However, it may happen that at some stage \(k\), the semidefinite relaxation \((21)\) may be infeasible because \(J^k(y)\) is infinite for some values of \(y \in Y_k(\tilde{x}_{k-1})\). This is because the feasible set \(K(\tilde{x}_{k-1})\) 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 \((4, 4.5)\).
bounds $0.4x_i^* \leq x_i \leq 1.6x_i^*$. In this case, $f_i^*$ is much closer to $f^*$ and we obtain the global minimum $f^*$ with ALGO 1 followed by the local minimization subroutine; see Table II. Importantly, in ALGO 1, and before running the local optimization subroutine, one ends up with a non feasible point $\bar{x}$. Moreover, we had to sometimes use the dichotomy procedure of Remark [1] because if $Y_k$ is large, one may have $K_y = \emptyset$ for $y$ in some open subintervals of $Y_k$.

Problem 7.2.2 has 13 linear constraints and 4 nonlinear constraints with bilinear terms. To handle the non-polynomial function $x_i^{0.5}$, one uses the lifting $u_2^2 = x_i$, $u_i \geq 0$, $i = 5,6$. Problem 7.2.6 has only 3 variables, 6 linear bound constraints, and one highly nonlinear constraint (and criterion). Here one uses the lifting $u^2x_2 = 1$, $u \geq 0$, to handle the term $x_2^{-1}$. 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 $x \mapsto x'Qx$ on $\{-1,1\}^n$, for some real symmetric matrix $Q \in \mathbb{R}^{n \times n}$. In this case, $Y_k = \{-1,1\}$ and the marginal constraint $L_k(x_k^2) = \gamma_k$ in (21) 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 $\lambda_0, \lambda_1$. Therefore after solving (21) one decides $\bar{x}_k = -1$ if $p_i(-1) < p_i(1)$ (i.e. if $\lambda_1 > 0$) and $\bar{x}_k = 1$ otherwise.

Recall that in ALGO 2 one first compute $\bar{x}_1$, then with $x_1$ fixed at the value $\bar{x}_1$, one computes $\bar{x}_2$, etc. until one finally computes $\bar{x}_n$, and get $\bar{x}$. In what we call the "max-gap" variant of ALGO 2, one first solves $n$ programs (4)- (5) with parameter $x_1$ to obtain an optimal solution $p_i(x_1) = \lambda_0^i + \lambda_1^i x_1$ of the dual (5), then with $x_2$ to obtain ($\lambda_0^1, \lambda_1^1$), etc. finally with $x_n$ to obtain ($\lambda_0^n, \lambda_1^n$). One then select $k$ such that $|\lambda_1^k| = \max_i |\lambda_1^i|$, and compute $\bar{x}_k$ 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 $\bar{x}_k$, one repeats the procedure for the $(n - 1)$-problem $P(\bar{x}_k)$, 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 $\varphi_k$ on $Y_k = \{-1,1\}$ is uniform (i.e. $\beta_1 = 0$ in (20)). Let $f_k^*$ denote the optimal value of the Shor's relaxation with famous Goemans and Williamson's 0.878 performance guarantee. Let $\rho$ denote the cost of the solution $x \in \{-1,1\}^n$ generated by the ALGO 2. In Table III we have reported the average relative error $(\rho - f_k^*)/|f_k^*|$, which as one may see, is comparable with the Goemans and Williamson (GW) ratio.

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 $K_y = \emptyset$ for $y$ in some open subinterval of $Y_k$ (proceeding by dichotomy on $Y_k$ is one possibility).

REFERENCES

[1] C.A. Floudas et al., Handook 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.

| Prob | n | m | $f^*$ | $r$ | 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%      |
| 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%        |
| 5.2.4 | 9  | 24| 750 | 1  | 750 | 0%        |
| 7.2.2 | 6 | 17| -0.3746 | -1 | -0.3746 | 0%        |
| 7.2.6 | 3 | 7 | -83.254 | -1 | -82.3775 | 1%        |

| $n$ | 20 | 30 | 40 |
|-----|----|----|----|
| $(\rho - f_i^*)/|f_i^*|$ | 10.3% | 12.3% | 12.5% |

TABLE II
ALGO 1 FOR NON CONVEX SET K

TABLE III
RELATIVE ERROR FOR MAXCUT