Solving Global Optimization Problems with Sparse Polynomials and Unbounded Semialgebraic Feasible Sets*

V. Jeyakumar[†], S. Kim [‡], G. M. Lee [§] and G. Li[¶]

June 6, 2014

Abstract

We propose a hierarchy of semidefinite programming (SDP) relaxations for polynomial optimization with sparse patterns over unbounded feasible sets. The convergence of the proposed SDP hierarchy is established for a class of polynomial optimization problems. This is done by deriving a new sum of squares sparse representation of positivity for a system of coercive polynomials over unbounded semialgebraic sets. We demonstrate that the proposed sparse SDP hierarchy can solve some classes of large scale polynomial optimization problems with unbounded feasible sets using the polynomial optimization solver SparsePOP developed by Waki et al. [24].

1 Introduction

The optimal value of a polynomial optimization over a compact semialgebraic set can be approximated as closely as desired by solving a hierarchy of semidefinite programming (SDP) relaxations and the convergence is finite generically under a mild assumption that requires the compactness of the feasible region (see [11, 14, 17]). It is known that the size of the SDP relaxations of the hierarchy, known now as the Lasserre hierarchy, rapidly grows as the number of variables and the relaxation order increase, preventing applications of the hierarchy to large scale polynomial optimization problems as the size of the SDP relaxations are too large to solve. A great deal of attention has recently been focused on reducing the size of these SDP relaxations. This has led to a sparse variant of the

*The first and the fourth author were partially supported by a grant from Australian Research Council. The second author was supported by NRF 2012-R1A1A2-038982. The third author was supported by Basic Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(NRF-2013R1A1A2005378)

[†]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: v.jeyakumar@unsw.edu.au

[‡]Department of Mathematics, Ewha W. University, 52 Ewhayeodae-gil, Sudaemoon-gu, Seoul, Korea 120-750, Email: skim@ewha.ac.kr

[§]Department of Applied Mathematics, Pukyong National University, Busan 608-737, Korea. E-mail: gmlee@pknu.ac.kr,

[¶]Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia. E-mail: g.li@unsw.edu.au

Lasserre hierarchy that allowed applications to various large scale polynomial optimization problems over compact semialgebraic feasible sets. (see also [8, 9, 23, 3, 7]).

More recently, the standard Lasserre hierarchy of SDP relaxations has been shown to extend to polynomial optimization over unbounded semialgebraic feasible sets via suitable modifications [5]. The purpose of this paper is to present a convergent *sparse SDP hierarchy* for solving polynomial optimization problems with sparse patterns and unbounded semialgebraic feasible sets, extending the unbounded version of the Lasserre hierarchy [5]. More specifically, we make the following contributions to Global polynomial optimization.

(1) We first establish a sparse version of the Putinar Positivstellensatze for coercive polynomials over unbounded semialgebraic sets which plays a key role in the construction and the proof of convergence of our sparse hierarchy of SDP relaxations. The Putinar Positivstellensatze for polynomials over compact semialgebraic sets that lead to the convergent Lasserre hierarchy can be found in (see [12, 9]).

(2) We then present a new sparse SDP hierarchy for solving polynomial optimization problems with unbounded feasible sets incorporating the objective function in the construction of quadratic modules that generate the sequence of SDP relaxations. This approach extends the relaxation scheme, developed for convex polynomial optimization over noncompact sets [6].

(3) By solving some numerical test problems, we illustrate that our sparse SDP hierarchy can easily be adapted with the current large scale polynomial optimization solver SparsePOP [24] to solve some classes of large scale polynomial optimization problems with unbounded feasible sets.

(4) We apply our SDP hierarchy to solve a class of sparse polynomial optimization problems with unbounded feasible sets and hidden coercivity.

The organization of the paper is as follows. In Section 2, we fix the notation and recall some basic facts on polynomial optimization. In Section 3, we provide the sparse version of the Putinar Positivstellensatze which applies to unbounded semialgebraic sets. In Section 4, we present our sparse SDP hierarchy for polynomial optimization with unbounded feasible sets and establish its convergence. In Section 5, we illustrate how our proposed scheme works by solving various large scale numerical test problems. Finally, in Section 6, we present an application of our hierarchy to a class of polynomial optimization problems with sparse patterns and hidden coercivity.

2 Preliminaries

Throughout this paper, $\mathbb{R}^n$ denotes the Euclidean space with dimension n. The inner product in $\mathbb{R}^n$ is defined by $\langle x, y\rangle := x^T y$ for all $x, y \in \mathbb{R}^n$. The non-negative orthant of $\mathbb{R}^n$ is denoted by $\mathbb{R}^n_+$ and is defined by $\mathbb{R}^n_+ := \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid x_i \geq 0\}$. The closed ball with center x and radius r is denoted by $\overline{\mathbb{B}}(x, r)$. We use $\mathbf{e}_n$ to denotes the vector in $\mathbb{R}^n$ whose elements are all one. Denote by $\mathbb{R}[\underline{x}]$ the ring of polynomials in

$x := (x_1, x_2, \dots, x_n)$ with real coefficients. The degree of a real polynomial f is denoted by $\deg f$. We say that a real polynomial $f \in \mathbb{R}[\underline{x}]$ is a *sum of squares* (SOS) polynomial if there exist real polynomials $f_j, j = 1, \dots, r$, such that $f = \sum_{j=1}^r f_j^2$. The set of all sum of squares real polynomials with variable x is denoted by $\Sigma^2[\underline{x}]$. The set of all sum of squares real polynomial with variable x and degree at most d is denoted by $\Sigma_d^2[\underline{x}]$. An important property of SOS polynomials is that checking a polynomial is sum of squares or not is equivalent to solving a semidefinite linear programming problem (see [11, 14, 17]).

Recall that a *quadratic module generated by polynomials* $-g_1, \dots, -g_m \in \mathbb{R}[\underline{x}]$ is defined as

$$\mathbf{M}(-g_1, \dots, -g_m) := \{\sigma_0 - \sigma_1 g_1 - \cdots - \sigma_m g_m \mid \sigma_i \in \Sigma^2[\underline{x}], i = 0, 1, \dots, m\}.$$

It is a subset of polynomials that are non-negative on the set $\{x \in \mathbb{R}^n \mid g_i(x) \leq 0, i = 1, \dots, m\}$ and possess a very nice certificate for this property.

The quadratic module $\mathbf{M}(-g_1, \dots, -g_m)$ is called *Archimedean* [14, 11] if there exists $p \in \mathbf{M}(-g_1, \dots, -g_m)$ such that $\{x : p(x) \geq 0\}$ is compact. When the quadratic module $\mathbf{M}(-g_1, \dots, -g_m)$ is compact, we have the following important characterization of positivity of a polynomial over a compact semialgebraic set.

Lemma 2.1. (Putinar positivstellensatz) [18] *Let f, g_j, $j = 1, \dots, m$, be real polynomials with $K := \{x : g_j(x) \leq 0, j = 1, \dots, m\} \neq \emptyset$. Suppose that $f(x) > 0$ for all $x \in K$ and $M(-g_1, \dots, -g_m)$ is Archimedean. Then, $f \in M(-g_1, \dots, -g_m)$.*

We now introduce a sparse version of Putinar positivestellensatz which was derived by Lasserre [12] and improved later on by Kojima et al [9]. Recall that, for a polynomial $f(x) = \sum_\alpha f_\alpha x^\alpha$ on $\mathbb{R}^n$ with degree d, the support of f is denoted by $\mathrm{supp} f$ and is defined by

$$\mathrm{supp} f = \{\alpha \in (\mathbb{N} \cup \{0\})^d : f_\alpha \neq 0\}.$$

Lemma 2.2. (Sparse version of Putinar positivstellensatz [12, 9]) *Let $f(x) = \sum_{l=1}^p f^l(x)$ and $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \dots, m$, be polynomials on $\mathbb{R}^n$ with degree d. Let I_l be a set of indexes such that $\mathrm{supp} f^l \subseteq I_l \subseteq \{1, \dots, n\}$, $l = 1, \dots, p$ and $\bigcup_{l=1}^p I_l = \{1, \dots, n\}$. Suppose that for each $j = 1, \dots, m$, $\mathrm{supp} g_j \subseteq I_l$ for some $l \in \{1, \dots, p\}$, and the following running intersection property holds: for each $l = 1, \dots, p-1$, there exists $s \leq l$ such that*

$$I_{l+1} \cap (\bigcup_{j=1}^{l} I_j) \subseteq I_s.$$

Let $K := \{x : g_j(x) \leq 0, j = 1, \dots, m\} \neq \emptyset$ and let $M(-g_1, \dots, -g_m)$ be Archimedean. If $f(x) > 0$ on K, then

$$f = \sum_{l=1}^{p} (\sigma_{0l} - \sum_{j=1}^{m} \sigma_{jl} g_j)$$

where σ_{jl}, $j = 0, 1, \dots, m$, are SOS polynomials with variables $\{x_i : i \in I_l\}$.

We note that, the assumption "$\mathrm{supp} g_j \subseteq I_l$ for some $l \in \{1, \dots, p\}$" and the running intersection property are automatically satisfied in the special case where $p = 1$ and

$I_1 = \{1, \ldots, n\}$. So, in this case, the sparse version of Putinar positivstellensatz reduces to Putinar positivstellensatz.

3 Sparse representations of positivity over unbounded sets

In this Section, we provide SOS representations of positivity of a class of nonconvex polynomials over unbounded semialgebraic sets. To do this, we first recall the definitions of coercive polynomials and strongly coercive polynomials.

Definition 3.1. (Coerciveness and Strong Coerciveness) *Let $f(x) = \sum_\alpha f_\alpha x^\alpha$ be a polynomial on $\mathbb{R}^n$ with degree d. Let $f(x) = \sum_{i=0}^d f_i(x)$ where each f_i is a homogeneous polynomial of degree i, $i = 0, 1, \ldots, d$. We say the polynomial f is*

- **coercive** *if $f(x) \to +\infty$ whenever $\|x\| \to +\infty$;*
- **s-strongly coercive** *for some $s \in \{1, \ldots, d\}$ if $f_s(x) > 0$ for all $x \neq 0$ and $f_i(x) \geq 0$ for all $s + 1 \leq i \leq d$ and $x \in \mathbb{R}^n$;*
- **strongly coercive** *if f is d-strongly coercive.*

It follows from the definition that a s-strongly coercive polynomial, $s = 1, \ldots, d$, must be coercive. On the other hand, the converse is not true. As an example, the 2-dimensional Rosenbrock function

$$f(x_1, x_2) = 1 + (x_2 - x_1^2)^2 + (1 - x_2)^2$$

is a coercive polynomial which is not s-strongly coercive for $s = 1, 2, 3, 4$. We also note that it was shown in [5] that the strong coercivity can be numerically checked by solving semidefinite programming problems. Furthermore, any polynomial of the form $\sum_{i=1}^n a_i x_i^d + \sum_{|\alpha| \leq d-1} h_\alpha x^\alpha$ where $|\alpha| = \sum_{i=1}^n \alpha_i$ with $\alpha = (\alpha_1, \ldots, \alpha_n) \in (\mathbb{N} \cup \{0\})^n$, $d \in 2\mathbb{N}$ and $a_i > 0$, $i = 1, \ldots, n$, is a strongly coercive polynomial.

The following proposition shows that a coercive polynomial is always level-bounded. Moreover, the corresponding bound for the lower level set can be computed using the coefficients of the underlying polynomial if the polynomial is assumed to be s-strongly coercive for some $s = 1, \ldots, d$.

Proposition 3.1. (Boundness of the lower level set via coercivity) *Let $f(x) = \sum_\alpha f_\alpha x^\alpha$ be a coercive polynomial on $\mathbb{R}^n$ with degree d. Let $f(x) = \sum_{i=0}^d f_i(x)$ where each f_i is a homogeneous polynomial with degree i, $i = 0, 1, \ldots, d$. Then, for each $c \in \mathbb{R}$, the lower level set $\{x : f(x) \leq c\}$ is a compact set. Furthermore, if f is s-strongly coercive for some $s = 1, \ldots, d$, then $\{x : f(x) \leq c\} \subseteq \overline{\mathbb{B}}(0, r)$ where*

$$r = \max\{1, \frac{c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha|}{\rho_s}\} \text{ and } \rho_s = \min\{f_s(x) : \|x\| = 1\}.$$

Proof. Fix any $c \in \mathbb{R}$. Then, the lower level set $\{x : f(x) \leq c\}$ is a compact set. To see this, we suppose on the contrary that there exists $\{x_n\}_{n=1}^\infty$, with $f(x_n) \leq c$ and $\{x_n\}$ is unbounded. By passing to subsequence if necessary, we may assume that $\|x_n\| \to +\infty$.

As f is a coercive polynomial, we must have $f(x_n) \to +\infty$. This contradicts the fact that $f(x_n) \leq c$ for all $n \in \mathbb{N}$.

To see the second assertion, we assume that f is s-strongly coercive for some $s = 1, \ldots, d$. Let $a \in \mathbb{R}^n$ be any point such that $f(a) \leq c$. Then,

$$\rho_s \|a\|^s \leq f_s(a) \leq \sum_{j=s}^{d} f_j(a) = f(a) - \sum_{j=0}^{s-1} f_j(a) \leq c - \sum_{j=0}^{s-1} f_j(a) \leq c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha|\, |a^\alpha|.$$

This gives us that either $\|a\| \leq 1$ or

$$\rho_s \|a\|^s \leq c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha|\, |a^\alpha| \leq c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha| \|a\|^{|\alpha|} \leq (c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha|) \|a\|^{s-1},$$

where the second inequality follows from the fact that $|x^\alpha| \leq \|x\|^{|\alpha|}$ for all $\|x\| \geq 1$ and the last inequality is from the fact that $\|x\|^{|\alpha|} \leq \|x\|^{s-1}$ for all $|\alpha| \leq s-1$ and $\|x\| \geq 1$. So, we have

$$\|a\| \leq \max\{1, \frac{c + \sum_{0 \leq |\alpha| \leq s-1} |f_\alpha|}{\rho_s}\},$$

and hence, the conclusion follows. $\square$

Corollary 3.1. *Let $f(x) = \sum_\alpha f_\alpha x^\alpha$ and $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \ldots, m$, be polynomials on $\mathbb{R}^n$ with degree d.*

(i) *If there exist $\mu_j \geq 0$, $j = 0, 1, \ldots, m$, such that $\mu_0 f + \sum_{j=1}^m \mu_j g_j$ is coercive, then, for each $c \in \mathbb{R}$, the set $\{x : g_j(x) \leq 0, j = 1, \ldots, m, f(x) \leq c\}$ is a compact set.*

(ii) *If $\mu_0 f + \sum_{j=1}^m \mu_j g_j$ is s-strongly coercive for some $s \in \{1, \ldots, d\}$, then*

$$\{x : g_j(x) \leq 0, j = 1, \ldots, m, f(x) \leq c\} \subseteq \overline{\mathbb{B}}((0, r),$$

where

$$r = \max\{1, \frac{\mu_0 c + \sum_{0 \leq |\alpha| \leq s-1} |\mu_0 f_\alpha + \sum_{j=1}^m \mu_j (g_j)_\alpha|}{\rho_s}\}$$

and

$$\rho_s = \min\{(\mu_0 f + \sum_{j=1}^m \mu_j g_j)_s(v) : \|v\| = 1\}.$$

Proof. Note that

$$\{x : g_j(x) \leq 0, j = 1, \ldots, m, f(x) \leq c\} \subseteq \{x : \mu_0 f(x) + \sum_{j=1}^m \mu_j g_i(x) \leq \mu_0 c\}.$$

The conclusion follows by applying the Proposition 3.1 with f replaced by $\mu_0 f + \sum_{j=1}^m \mu_j g_i$. $\square$

We now present a sparse representation result for positivity of polynomials over a unbounded semialgebraic set. The proof of this result makes use of Lemma 2.2, the sparse Putinar positivestellensatz, and Proposition 3.1.

Theorem 3.1. (Sparse representation for positivity over an unbounded semialgebraic set) *Let $f(x) = \sum_{l=1}^{p} f^l(x)$ and $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \ldots, m$, be polynomials on $\mathbb{R}^n$ with degree d. Let I_l be a set of indexes such that* $\mathrm{supp} f^l \subseteq I_l \subseteq \{1, \ldots, n\}$, *$l = 1, \ldots, p$ and $\bigcup_{l=1}^{p} I_l = \{1, \ldots, n\}$. Let $K = \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$ and $x^0 \in K$. Let $c > \max_{1 \leq l \leq p}\{f^l(x^0)\}$.Suppose that for each $j = 1, \ldots, m$,* $\mathrm{supp} g_j \subseteq I_l$ *for some $l \in \{1, \ldots, p\}$, and the following running intersection property holds: for each $l = 1, \ldots, p-1$, there exists $s \leq l$ such that*

$$I_{l+1} \cap (\bigcup_{j=1}^{l} I_j) \subseteq I_s.$$

Assume that for each $l = 1, \ldots, p$, there exist $\mu_{0l} \geq 0$, $l = 1, \ldots, p$ $\mu_j \geq 0$, $j = 1, \ldots, m$, such that $\sum_{l=1}^{p} \mu_{0l} f^l + \sum_{j=1}^{m} \mu_j g_j$ is coercive, and $f > 0$ over K. Then, there exist sum-of squares polynomials $\sigma_{0l}, \ldots, \sigma_{ml}, \bar{\sigma}_l$ with variables $\{x_i : i \in I_l\}$, $l = 1, \ldots, p$, such that

$$f = \sum_{l=1}^{p} \big(\sigma_{0l} - \sum_{j=1}^{m} \sigma_{jl} g_{jl} + \bar{\sigma}_l (c - f^l)\big).$$

Proof. It follows from Proposition 3.1 that $\{x : \sum_{l=1}^{p} \mu_{0l}(c - f^l(x)) + \sum_{j=1}^{m} \mu_j(-g_j)(x) \geq 0\}$ is compact. So, by definition, we see that $\mathbf{M}(-g_1, \ldots, -g_m, c - f^1, \ldots, c - f^p)$ is Archimedean. Note that $f > 0$ over $\hat{K}$ where

$$\hat{K} = \{x : g_j(x) \leq 0, j = 1, \ldots, m, f^l(x) - c \leq 0, l = 1, \ldots, p\}.$$

Hence, by Lemma 2.2, we obtain that

$$f = \sum_{l=1}^{p} \big(\sigma_{0l} - \sum_{j=1}^{m} \sigma_{jl} g_j + \bar{\sigma}_l (c - f^l)\big),$$

for some sum-of squares polynomials $\sigma_{0l}, \ldots, \sigma_{ml}, \bar{\sigma}_l$ with variables $\{x_i : i \in I_l\}$, $l = 1, \ldots, p$. □

Remark 3.1. *It is worth noting that, if a polynomial f on $\mathbb{R}^n$ is convex in the sense that $\nabla^2 f(x)$ is positive semi-definite for all $x \in \mathbb{R}^n$ and there exists $x^* \in \mathbb{R}^n$ such that $\nabla^2 f(x^*)$ is positive definite, then it is coercive (for example see [6]). Therefore, our coercive assumption of Theorem 3.1 that, "there exist $\mu_{0l} \geq 0$, $l = 1, \ldots, p$ $\mu_j \geq 0$, $j = 1, \ldots, m$, such that $\sum_{l=1}^{p} \mu_{0l} f^l + \sum_{j=1}^{m} \mu_j g_j$ is coercive" is satisfied, if one of the polynomials f^l, g_j, $l = 1, \ldots, p$ and $j = 1, \ldots, m$, is convex and has a positive definite Hessian at some point $x^* \in \mathbb{R}^n$.*

As a corollary, we obtain the dense representation of positivity of a polynomial over a noncompact semialgebraic set given in [5].

Corollary 3.2. (Representation of positivity over an unbounded semialgebraic set) *Let $f(x) = \sum_\alpha f_\alpha x^\alpha$ and $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \ldots, m$, be polynomials on $\mathbb{R}^n$ with degree d. Let $K = \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$, $x^0 \in K$ and $c > f(x^0)$. Suppose* *that there exist* $\mu_j \geq 0$, $j = 0, 1, \ldots, m$, *such that* $\mu_0 f + \sum_{j=1}^m \mu_j g_j$ *is coercive , and* $f > 0$ *over* K. *Then, there exist sum-of squares polynomial* $\sigma_0, \ldots, \sigma_m, \bar{\sigma}$ *such that*

$$f = \sigma_0 - \sum_{i=1}^{m} \sigma_j g_j + \bar{\sigma}(c - f).$$

Proof. Let $p = 1$ and $I_1 = \{1, \ldots, n\}$. Note that the assumption "for each $j = 1, \ldots, m$, $\mathrm{supp} g_j \subseteq I_l$" for some $l \in \{1, \ldots, p\}$, and the running intersection property holds automatically in this case. Then, the conclusion follows from the preceding theorem. □

4 A sparse hierarchy for optimization over unbounded sets

Consider the polynomial optimization problem

$$\begin{array}{lll} (P) & \min & \sum_{l=1}^{p} f^l(x) \\ & \text{s.t.} & g_j(x) \leq 0, j = 1, \ldots, m, \end{array}$$

where f^l, g_j are (nonconvex) polynomials on $\mathbb{R}^n$, $l = 1, \ldots, p$ and $j = 1, \ldots, m$. Let the feasible set be denoted by K, that is, $K = \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$.

Let $f(x) = \sum_{l=1}^p f^l(x)$ and $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \ldots, m$, be polynomials on $\mathbb{R}^n$ with degree d. Let I_l be a set of indices such that $\mathrm{supp} f^l \subseteq I_l \subseteq \{1, \ldots, n\}$, $l = 1, \ldots, p$ and $\bigcup_{l=1}^p I_l = \{1, \ldots, n\}$. Let $x^0 \in \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$ and let c be a number such that $c > \max_{1 \leq l \leq p}\{f^l(x^0)\}$. For each integer k, we define the truncated sparse version of the quadratic module $\bar{\mathbf{M}}_k$ by

$$\begin{aligned} \bar{\mathbf{M}}_k := \{ \quad & \sum_{l=1}^{p} \left(\sigma_{0l} + \sum_{j \in I_l} \sigma_{jl} g_j + \bar{\sigma}_l (c - f^l) \right) \mid \sigma_{0l}, \sigma_{jl}, \bar{\sigma}_l \in \Sigma^2[\underline{x_l}], \\ & \deg \sigma_{0l} \leq 2k, \deg \sigma_{jl} g_j \leq 2k, \deg \bar{\sigma}_l (c - f^l) \leq 2k\}, \end{aligned}$$

where $\Sigma^2[\underline{x_l}]$, $l = 1, \ldots, p$, denotes the set of all SOS polynomials with variable $\{x_i : i \in I_l\}$.

Consider the following relaxation problem

$$\bar{f}_k^* := \sup\{\mu \in \mathbb{R} \mid f - \mu \in \bar{\mathbf{M}}_k\}. \tag{4.1}$$

By construction, $\bar{f}_k^* \leq \bar{f}_{k+1}^* \leq \cdots \leq \min(P)$. Note that, if we set $\bar{\sigma}_l \equiv 0$, $l = 1, \ldots, p$, then the hierarchy (4.1) reduces to the known sparse SDP hierarchy proposed in [8, 9].

Theorem 4.1. (Convergence of sparse SDP hierarchy) *Let* $f(x) = \sum_{l=1}^p f^l(x)$ *and* $g_j(x) = \sum_\alpha (g_j)_\alpha x^\alpha$, $j = 1, \ldots, m$, *be polynomials on* $\mathbb{R}^n$ *with degree* d. *Let* I_l *be a set of indexes such that* $\mathrm{supp} f^l \subseteq I_l \subseteq \{1, \ldots, n\}$, $l = 1, \ldots, p$ *and* $\bigcup_{l=1}^p I_l = \{1, \ldots, n\}$. *Consider the SDP hierarchy (4.1) and denote its optimal value by* $\bar{f}_k^*$. *Let* $K = \{x : g_j(x) \leq 0, j = 1, \ldots, m\}$, $x^0 \in K$ *and* $c > \max_{1 \leq l \leq p}\{f^l(x^0)\}$. *Suppose that for each*

$j = 1, \ldots, m$, $\mathrm{supp} g_j \subseteq I_l$ for some $l \in \{1, \ldots, p\}$, and the following running intersection property holds: for each $l = 1, \ldots, p-1$, there exists $s \leq l$ such that

$$I_{l+1} \cap (\bigcup_{j=1}^{l} I_j) \subseteq I_s.$$

If, for each $l = 1, \ldots, p$, there exist $\mu_{0l} \geq 0$, $l = 1, \ldots, p$, $\mu_j \geq 0$, $j = 1, \ldots, m$, such that $\sum_{l=1}^{p} \mu_{0l} f^l + \sum_{j=1}^{m} \mu_j g_j$ is coercive, then $\lim_{k\to\infty} \bar{f}_k^* = \min(P)$.

Proof. Let $\epsilon > 0$. Then, we have $f - \min(P) + \epsilon > 0$ over the feasible set K. As there exist $\mu_j \geq 0$, $j = 0, 1, \ldots, m$, such that $\mu_0 f + \sum_{j=1}^{m} \mu_j g_j$ is coercive, we see that $\mu_0(f - \min(P) + \epsilon) + \sum_{j=1}^{m} \mu_j g_j$ is also coercive. Thus, the sparse positivity representation result (3.1) implies that there exist sum-of squares polynomial $\sigma_{0l}, \ldots, \sigma_{jl}, j = 1, \ldots, m, \bar{\sigma}_l \in \mathbb{R}[\underline{x_l}]$, $l = 1, \ldots, p$, such that

$$f - \min(P) + \epsilon = \sum_{l=1}^{p} \big(\sigma_{0l} - \sum_{j=1}^{m} \sigma_{jl} g_j + \bar{\sigma}_l (c - f^l)\big).$$

Thus, for each $\epsilon > 0$, there exists $k_0 \in \mathbb{N}$ such that $\bar{f}_k^* \geq \min(P) - \epsilon$. On the other hand, from the construction of the hierarchy, we see that $\bar{f}_k^* \leq \bar{f}_{k+1}^* \leq \cdots \leq \min(P)$. Therefore, the conclusion follows. $\square$

In the special case where $p = 1$ and $I_1 = \{1, \ldots, n\}$, the optimization problem (P) reduces to

$$\begin{array}{cl} \min_{x\in\mathbb{R}^n} & f(x) \\ \text{s.t.} & g_j(x) \leq 0, j = 1, \ldots, m. \end{array}$$

In this case, the sparse truncated quadratic module $\bar{\mathbf{M}}_k$ reduces to the truncated quadratic module $\mathbf{M}_k$ generated by the polynomials $c - f$ and $-g_1, \ldots, -g_m$ given by

$$\begin{aligned} \mathbf{M}_k \; := \; \{\sigma_0 + \sum_{j=1}^{m} \sigma_j g_j + \bar{\sigma}(c - f) \quad | \quad & \sigma_0, \sigma, \bar{\sigma} \in \Sigma^2[\underline{x}] \subset \mathbb{R}[\underline{x}], \\ & \deg\, \sigma_0 \leq 2k,\ \deg\, \sigma_j g_j \leq 2k,\ \text{and}\ \deg \bar{\sigma}(c - f) \leq 2k\}, \end{aligned}$$

and the corresponding relaxation problem collapses to

$$f_k^* := \sup\{\mu \in \mathbb{R} \mid f - \mu \in \mathbf{M}_k\}. \tag{4.2}$$

So, Theorem 4.1 collapses to the convergence result of the dense SDP hierarchy for polynomial optimization problem with noncompact sets proposed in [5].

5 Numerical experiments

In this Section, we show the effectiveness of the proposed sparse SDP hierarchy in Section 4 by solving some numerical test problems with unbounded feasible sets. All the numerical tests were conducted on a computer, with a 2.8 GHz Intel Core i7 and 8 GB RAM, equipped with Matlab 7.14 (R2012a).

The purpose of the numerical experiments is to illustrate how our proposed sparse SDP hierarchy works for solving polynomial optimization problems with unbounded feasible sets. Therefore, we first select some known test problems which are coercive and test them by minimizing them over unbounded feasible sets. The selected test problems are:

(1) a nonconvex quadratic programming problem with an unbounded feasible set, in which Lasserre's hierarchy is known to fail;

(2) the Rosenbrock function over the nonnegative orthant;

(3) the Chain-wood function over the nonnegative orthant;

For the numerical tests, let the objective function $f(x)$ take the form $f(x) = \sum_{l=1}^{p} f^l(x)$. We added additional constraints $f^l(x) \leq c$, $l = 1, \ldots, q$, to the test problems with unbounded feasible sets for the proposed sparse SDP hierarchy, where c is appropriately chosen. We note that the resulting test problems are different from the Rosenbrock function and the Chain-wood function solved in [23]. We then used a Matlab software SparsePOP [24] to solve the problems. SparsePOP can solve polynomial optimization problems exploiting the sparsity described in Section 4 by setting the parameter *sparseSW = 1*, and can also implement Lasserre's hierarchy by setting the parameter *sparseSW = 0*. In addition, a parameter, called the relaxation order, can be chosen in SparsePOP, depending on the degree of the polynomial optimization problem. The larger value for the relaxation order is used, the better approximation to the optimal value of the polynomial optimization problem can be expected.

In SparsePOP, the accuracy of an obtained objective value is computed by

$$\text{Relative objective error (Rel.Er)} \;=\; \frac{\text{POP.objValue - SDP.objValue}}{\max\{1, |\text{POP.objValue}|\}},$$

where POP.objValue means the value of the objective function of a polynomial optimization problem computed using a candidate for the optimal solution, and SDP.objValue the value of the objective function of the SDP relaxation of the polynomial optimization problem. Moreover, CPU time reported in the subsequent discussion is measured in seconds. For details, we refer to [24].

5.1 A 2-dimensional QP with an unbounded feasible set

Consider the following nonconvex quadratic optimization problem:

$$\begin{array}{lll} (QP) & \min_{(x_1,x_2)\in\mathbb{R}^2} & x_1^2 + x_2^2 \\ & \text{s.t.} & x_2^2 - 1 \geq 0 \\ & & x_1^2 - Mx_1x_2 - 1 \geq 0 \\ & & x_1^2 + Mx_1x_2 - 1 \geq 0. \end{array} \tag{5.3}$$

It was shown in [15, 13] that the global minimizers are $(\pm\frac{M+\sqrt{M^2+4}}{2}, \pm 1)$ and its global minimum is $1+\frac{\sqrt{2M^2+4+2M\sqrt{M^2+4}}}{4}$. For instance, if $M=5$, then, the global minimizers are $(\pm 5.1926, \pm 1)$ and its global minimum is 27.9629.

Let $M=5$. It was shown in [15] that Lasserre's hierarchy only provides a lower bound 2, no matter how large the relaxation order k is chosen. As explained in [15], the main reason why the Lasserre's hierarchy fails to achieve the global minimum of this quadratic problem is that the feasible set of this nonconvex quadratic problem is unbounded.

Notice that the objective $f(x_1,x_2) := x_1^2 + x_2^2$ is coercive. Thus, we can apply the proposed sparse hierarchy. Since the problem has only two variables without sparsity, we test the problem using our dense SDP hierarchy (4.2) with $c=40$. Note that $c > f(6,1)$ and the point $(6,1)$ is feasible for this quadratic optimization problem.

As shown in Table 5.1, by running the sparsePOP with sparseSW=0 and the relaxation order 4, the relaxation problem in the dense hierarchy (4.2) solves the original problem and returns a good approximation to the true global minimizer $(5.1296, 1.000)$. Table 5.1 summarizes the optimal values for the dense SDP hierarchy (4.1) with relaxation order $k=2,3,4$, which illustrates the effectiveness of our approach.

	RelaxOrder	c	Optimal val.	Optimal sol.	Rel.Er	CPU
DenseHierarchy (4.2)	2	40	1.1729	-	3.9e+0	0.27
DenseHierarchy (4.2)	3	40	14.2398	-	5.6e-5	0.31
DenseHierarchy (4.2)	4	40	27.9629	(5.1296,1.000)	2.6e-5	0.39

Table 5.1: Numerical test on problem (QP2).

5.2 The Rosenbrock function over nonnegative orthant

The Rosenbrock function is described as

$$f_R(x_1,\ldots,x_n) = 1 + \sum_{i=2}^{n} \left((x_i - x_{i-1}^2)^2 + (1-x_i)^2\right),\ n \geq 2.$$

Clearly, f_R is a SOS polynomial, and is coercive. We add constraints to the Rosenbrock function to have a polynomial optimization problem with an unbounded region as follows:

$$\begin{array}{lll} (EP_R) & \min_{x\in\mathbb{R}^n} & f_R(x) \\ & \text{s.t.} & x_i \geq 0, \quad i=1,\ldots,n. \end{array}$$

It can be easily verified that this problem has a unique global minimizer $\mathbf{e}_n := (\underbrace{1,\ldots,1}_{n})$.

Let $I_l = \{l, l+1\}$, $l=1,\ldots,n-1$ and $g_j(x) = x_j$, $j=1,\ldots,n$. Then, the assumptions in Theorem 4.1 are satisfied as

$$\text{supp} g_j = \{j\} \subseteq \left\{ \begin{array}{lll} I_j, & \text{if} & j=1,\ldots,n-1, \\ I_{n-1}, & \text{if} & j=n, \end{array} \right.$$

and for each $l = 1, \dots, n-1$,

$$I_{l+1} \cap (\bigcup_{j=1}^{l} I_j) = \{l\} \subseteq I_l.$$

Therefore, according to Theorem 4.1, the optimal value of the proposed sparse SDP hierarchy (4.1) converges to the optimal value of the global minimum of (EP_R).

We now test our proposed sparse SDP hierarchy (4.1) using the global optimization problem (EP_R) for different dimension n with $c = 2$. From Table 5.2, we see that for $n = 10000$ or $n = 20000$, the sparsePOP returns an accurate solution for the relaxation order 2 in the proposed sparse hierarchy Sparse SDP hierarchy (4.1).

	n	RelaxOrder	c	Optimal sol.	Rel.Er	CPU
SparseHierarchy (4.1)	500	2	2	$\mathbf{e}_{500}$	4.1e-5	81.8
SparseHierarchy (4.1)	5000	2	2	$\mathbf{e}_{5000}$	1.0e-3	148.5
SparseHierarchy (4.1)	10000	2	2	$\mathbf{e}_{10000}$	2.1e-3	308.5
SparseHierarchy (4.1)	20000	2	2	$\mathbf{e}_{20000}$	4.3e-3	772.2

Table 5.2: Numerical tests on the problem (EP_R).

When the dimension n of the problem is large, directly applying the dense SDP hierarchy proposed in [5] without exploiting the sparsity leads to very large SDP problems which cannot be handled by a SDP solver such as SeDuMi [22] . Indeed, we confirm in our numerical computation that the dense SDP hierarchy proposed in [5] can only be used to solve (EP_R) up to dimension 20. The larger problems than dimension 20 resulted in out-of-memory error.

5.3 Chained-wood function over nonnegative orthant

Let $n \in 4\mathbb{N}$ where $\mathbb{N}$ denotes the set of integers. Consider the Chained-wood function given by

$$\begin{aligned} f_C(x_1, \dots, x_n) &= 1 + \sum_{i \in J} \big((x_{i+1} - x_i^2)^2 + (1 - x_i)^2 + 90(x_{i+3}^2 - x_{i+2})^2 + (x_{i+2} - 1)^2 \\ &\quad + 10(x_{i+1} + x_{i+3} - 2)^2 + \frac{1}{10}(x_{i+1} - x_{i+3})^2 \big), \end{aligned}$$

where $J = \{1, 3, 5, \dots, n-3\}$. The function f_C is a SOS polynomial, and is coercive. Adding nonnegative constraints for the variables $x_i,\ i = 1, \dots, n$ results in the following polynomial optimization problem:

$$\begin{array}{lll} (EP_C) & \min_{x \in \mathbb{R}^n} & f_C(x) \\ & \text{s.t.} & x_i \geq 0, \quad i = 1, \dots, n. \end{array}$$

Clearly, the feasible set of this polynomial optimization is unbounded, and this problem has a unique global minimizer $\mathbf{e}_n := (\underbrace{1, \dots, 1}_{n})$.

Note that f_C can be equivalently rewritten as

$$\begin{aligned} f_C(x_1,\ldots,x_n) &= 1+\sum_{l=1}^{\frac{n}{2}-1}\big((x_{2l}-x_{2l-1}^2)^2+(1-x_{2l-1})^2+90(x_{2l+2}^2-x_{2l+1})^2+(x_{2l+1}-1)^2 \\ &\quad +10(x_{2l}+x_{2l+2}-2)^2+\frac{1}{10}(x_{2l}-x_{2l+2})^2\big). \end{aligned}$$

For each $l=1,\ldots,\frac{n}{2}-1$, let $I_l=\{2l-1,2l,2l+1,2l+2\}$ and $g_j(x)=x_j$, $j=1,\ldots,n$. Then, the assumptions in Theorem 4.1 are satisfied as $\mathrm{supp} g_j=\{j\}\subseteq I_{[\frac{j}{2}]}$, where $[\frac{j}{2}]$ is the largest integer that is smaller than $\frac{j}{2}$, and for each $l=1,\ldots,n-1$,

$$I_{l+1}\cap(\bigcup_{j=1}^{l} I_j)=\{2l+1,2l+2\}\subseteq I_l.$$

Therefore, Theorem 4.1 implies that the optimal value of the proposed sparse SDP hierarchy (4.1) converges to the optimal value of the global minimum of (EP_R).

We now test our proposed sparse SDP hierarchy (4.1) on (EP_C) for different values of n with $c=n+1$. From Table 5.3, we see that for dimension $n=10000$ or $n=20000$, the sparsePOP with the relaxation order =2 returns an accurate solution in the proposed sparse hierarchy Sparse SDP hierarchy (4.1).

	n	RelaxOrder	c	Optimal sol.	Rel.Er	CPU
SparseHierarchy (4.1)	1000	2	1001	$(\mathbf{e}_{1000})$	2.4e-4	38.9
SparseHierarchy (4.1)	5000	2	5001	$(\mathbf{e}_{5000})$	1.2e-3	175.2
SparseHierarchy (4.1)	10000	2	10001	$(\mathbf{e}_{10000})$	2.4e-3	392.3
SparseHierarchy (4.1)	20000	2	20001	$(\mathbf{e}_{20000})$	4.8e-3	1049.8

Table 5.3: Numerical tests on the problem (EP_C)

The numerical experiment using the dense SDP hierarchy proposed in [5] could solve (EP_C) up to only $n=10$. We observe again that much larger problems can be solved by the sparse SDP hierarchy.

6 A class of sparse problems with hidden coercivity

Consider the following polynomial optimization problem:

$$\begin{aligned} (P_0) \quad & \min_{(x^l,w^l)\in\mathbb{R}^{n_l}\times\mathbb{R}^{n_l},l=1,\ldots,q} && \sum_{l=1}^{q}\sum_{i\in I_l}(w_i^l)^2 \\ & \text{s.t.} && g_j^l(x^l)\le 0,\ j=1,\ldots,m,\ l=1,\ldots,q \\ & && (1-w_i^l)x_i^l=0,\ i\in I_l, l=1,\ldots,q, \\ & && \|x^l\|_2^2\le M^l,\ l=1,\ldots,q, \end{aligned}$$

where $I_l \subseteq \{1,\ldots,n\}$, $l = 1,\ldots,q$, $\bigcup_{l=1}^q I_l = \{1,\ldots,n\}$, $n_l = |I_l|$, that is, n_l is the cardinality of I_l, $l = 1,\ldots,q$, $I_l = \{i_1,\ldots,i_{n_l}\}$ and g_j^l are polynomials, $j = 1,\ldots,m$, $l = 1,\ldots,q$.

In Remark 6.1 we show the link between (P_0) and the problem of finding the solution with the least number of nonzero components of a system of polynomial inequalities. Problems of this kind arise in signal processing and statistics [1, 2].

We now introduce a SDP hierarchy for problem (P_0) using the results in Section 3. It is worth noting that, problem (P_0) is a minimization problem with variables $(x,w) \in \mathbb{R}^n \times \mathbb{R}^n$. Thus, the feasible set of (P_0) can be unbounded in general (for example, simply take $g_j \equiv 0$. Then, $(w,0)$ is feasible for problem (P_0) for any $w \in \mathbb{R}^n$).

Let $f(x,w) = \sum_{l=1}^q f_l(x^l,w^l)$, where $f_l(x^l,w^l) = \sum_{i\in I_l}(w_i^l)^2$ and $x = (x^1,\ldots,x^q) \in \mathbb{R}^{\sum_{l=1}^q n_l}$ and $w = (w^1,\ldots,w^q) \in \mathbb{R}^{\sum_{l=1}^q n_l}$. For each $l = 1,\ldots,q$, define $\bar g_{il}(x^l,w^l) = (1-w_i^l)x_i^l$, $i \in I_l$ and $G_l(x^l,w^l) := \|x^l\|_2^2 - M^l$. Let c be a number such that $c > n^l$, $l = 1,\ldots,q$. For each integer k, we define the following sparse truncated quadratic module $\hat{\mathbf{M}}_k$ generated by the polynomials $c - f^l$, $-g_j^l$, $j = 1,\ldots,m$, $\bar g_{il}$, $i \in I_l$ and G_l, $l = 1,\ldots,q$, as follows:

$$
\begin{aligned}
\hat{\mathbf{M}}_k \ := \ & \Bigg\{ \sum_{l=1}^q \Bigg(\sigma_{0l} + \sum_{j=1}^m \sigma_{jl} g_j^l + \sum_{i=1}^{I_l} h_{il}\bar g_{il} + \sigma_l(-G_l) + \bar\sigma_l(c - f^l)\Bigg) \mid \sigma_{0l},\sigma_{il},\sigma_l,\bar\sigma_l \in \Sigma^2[\underline{x^l},\underline{w^l}], \\
& h_{il} \in \mathbb{R}[\underline{x^l},\underline{w^l}],\ \deg \sigma_{0l} \le 2k,\ \deg \sigma_{jl} g_{jl} \le 2k\ \deg h_{il}\bar g_{il} \le 2k, \text{ and} \\
& \deg \sigma_l G_l \le 2k, \deg \bar\sigma(c - f^l) \le 2k \Bigg\}.
\end{aligned}
$$

Consider the following relaxation problem

$$
f_k^* := \sup\{\mu \in \mathbb{R} \mid f - \mu \in \hat{\mathbf{M}}_k\}. \tag{6.4}
$$

Then, one can show that $\lim_{k\to\infty} f_k^* = \min(P_0)$ if the running intersection property holds.

Proposition 6.1. (Convergence of the sparse SDP hierarchy value for (P_0)) *For problem (P_0) and the SDP hierarchy (6.4), assume that the running intersection property that, for each $l = 1,\ldots,q-1$, there exists $s \le l$ such that $I_{l+1} \cap (\bigcup_{j=1}^l I_j) \subseteq I_s$, holds. Then,* $\lim_{k\to\infty} f_k^* = \min(P_0)$.

Proof. Note that the problem (P_0) can be equivalently rewritten as

$$
\begin{aligned}
\min_{(x^l,w^l)\in\mathbb{R}^{|I_l|}\times\mathbb{R}^{|I_l|}} \quad & \sum_{l=1}^q f^l(x^l,w^l) = \sum_{l=1}^q \sum_{i\in I_l} (w_i^l)^2 \\
\text{s.t.} \quad & g_j^l(x^l) \le 0,\ j = 1,\ldots,m, l = 1,\ldots,q \\
& \bar g_{il}(x^l,w^l) := (1-w_i^l)x_i^l \le 0,\ i \in I_l, l = 1,\ldots,q \\
& -\bar g_{il}(x^l,w^l) = -(1-w_i^l)x_i^l \le 0,\ i \in I_l, l = 1,\ldots,q \\
& G_l(x^l,w^l) := \|x^l\|_2^2 - M^l \le 0.
\end{aligned}
$$

As $\bigcup_{l=1}^p I_l = \{1,\ldots,n\}$, $\sum_{l=1}^p f^l(x^l,w^l) + \sum_{l=1}^p G_l(x^l,w^l)$ is strongly coercive. So, the assumptions in Theorem 4.1 are satisfied. Therefore, Theorem 4.1 implies that $\lim_{k\to\infty} f_k^* = \min(P_0)$. $\square$

We now illustrate the SDP hierarchy using a numerical example. For the numerical test on the effectiveness of the proposed SDP hierarchy, we let $q \in \mathbb{N}$. For each $l = 1, \ldots, q$, we generate a 3-by-4 matrix A_l containing random values drawn from the uniform distribution on the interval $[0, 1]$ and define $b_l = 4A_l(:, 1)$, where $A_l(:, 1)$ denotes the first column of A_l. Consider the problem:

$$
\begin{array}{lll}
(EP) & \min\limits_{x=(x^1,\ldots,x^q),x^l\in\mathbb{R}^4,l=1,\ldots,q} & \sum\limits_{l=1}^{q} \|x^l\|_0 \\
 & \text{s.t.} & A_l x^l = b_l,\ l = 1, \ldots, q \\
 & & \|x^l\|_2^2 \leq 100,\ l = 1, \ldots, q.
\end{array}
$$

It is not hard to see that an optimal solution of (EP) is $x^* := [x^{1*}, \ldots, x^{q*}]$ with $x^{l*} = [4, 0, 0, 0]$, $l = 1, \ldots, q$ with optimal value q.

Let $I_l = \{4l - 3, 4l - 2, 4l - 1, 4l\}$, $l = 1, \ldots, q$. Denote $x^l = (x_i^l)_{i\in I_l} \in \mathbb{R}^4$ and $w^l = (w_i^l)_{i\in I_l} \in \mathbb{R}^4$, $l = 1, \ldots, q$. In Remark 6.1, we see that problem (EP) is equivalent to

$$
\begin{array}{ll}
\min\limits_{\substack{x=(x^1,\ldots,x^q)\in\mathbb{R}^{4q} \\ w=(w^1,\ldots,w^q)\in\mathbb{R}^{4q}}} & \sum\limits_{l=1}^{q} \sum\limits_{i\in I_l} (w_i^l)^2 \\
\text{s.t.} & A_l x^l = b_l,\ l = 1, \ldots, q \\
 & (1 - w_i^l) x_i^l = 0,\ i \in I_l,\ l = 1, \ldots, q, \\
 & \|x^l\|_2^2 \leq 100,\ l = 1, \ldots, q.
\end{array}
$$

Clearly, the running intersection property holds. So, the optimal value of the proposed sparse SDP hierarchy (6.4) converges to the optimal value of (EP).

As shown in Table 6.1, for $q = 200$ and $q = 1000$, in our numerical experiment, sparsePOP with the relaxation order 2, $c = 5$ and $M = 100$, returns an accurate optimal value and an accurate optimal solution x^* in the proposed Sparse SDP hierarchy (6.4).

We also notice that the numerical experiment using the dense SDP hierarchy proposed in [5] could solve the above problem with number of blocks q up to only 5. We observe again that much larger problems can be solved by the sparse SDP hierarchy.

Problem	q, n	c	M	Optimal val.	Rel.Er	CPU
SparseHierarchy (6.4)	200, 1600	5	100	200	3.3e-9	173.9
SparseHierarchy (6.4)	1000, 8000	5	100	1000	1.1e-5	470.2

Table 6.1: Numerical tests on the polynomial optimization problem with hidden coercivity. n is the number of variables. The relaxation order 2 was used.

Finally, we note that one of the popular methods to find an approximate solution of problems (EP) is to replace the l_0 norm by the l_1 norm, and consider the following convex programming problem

$$(REP) \qquad \min_{x=(x^1,\ldots,x^q),x^l\in\mathbb{R}^4,l=1,\ldots,q} \quad \sum_{l=1}^{q} \|x^l\|_1$$
$$\text{s.t.} \qquad A_l x^l = b_l, \; l = 1,\ldots,q$$
$$\|x^l\|_2^2 \le 100, \; l = 1,\ldots,q,$$

where $\|\cdot\|_1$ is the l_1 norm defined by $\|x\|_1 = \sum_{i=1}^m |x_i|$ with $x = (x_1,\ldots,x_m) \in \mathbb{R}^m$. Solving this convex optimization problem provides an approximation of the solution of the problem (EP). This method is often referred as the l_1 heuristics [1] and is widely used, because problem (REP) is a convex problem which can be efficiently solved. On the other hand, problem (EP) is often hard to solve due to the nonconvexity and noncontinuity of the seminorm $\|\cdot\|_0$.

In general, the l_1 heuristics may not be able to provide an accurate solution for the problem (EP). For example, consider problem (EP) where $q = 200$,

$$A_l = A := \begin{pmatrix} 2 & -1 & 30 & 3 \\ 3 & 3 & 44 & 2 \\ -2 & 7 & -40 & -6 \end{pmatrix}, \; b_l = b := 4A(:,1) = \begin{pmatrix} 8 \\ 12 \\ -8 \end{pmatrix} \text{ for all } l = 1,\ldots,200.$$

Then, the true solution of this problem is $x^* := [x^{1*},\ldots,x^{q*}]$ with $x^{l*} = [4,0,0,0]$, $l = 1,\ldots,200$. Solving the corresponding l_1 heuristics (REP) gives us the solution $z^* = [z^{1*},\ldots,z^{200*}]$ with $z^{l*} = [0, 1.2571, 0.0857, 2.2286]$, $l = 1,\ldots,200$. Interestingly, the support of the solution z^* found by the l_1 heuristics is the complement of the true optimal solution and z^* is clearly not a solution of (EP) (as the corresponding objective value at z^* is 600 which is much larger than the true global optimal value 200). In contrast, in our numerical experiment, sparsePOP with the relaxation order 2 and $c = 5$, returns an accurate optimal value 200 and an accurate optimal solution x^*, in the proposed Sparse SDP hierarchy (6.4), as shown in Table 6.2.

Problem	q, n	c	M	Optimal val.	Rel.Er	CPU
SparseHierarchy (6.4)	200, 1600	5	100	200	5.9e-6	107.6
SparseHierarchy (6.4)	1000, 8000	5	100	1000	1.8e-6	493.5

Table 6.2: Numerical tests on the polynomial optimization problem with hidden coercivity. n is the number of variables. The relaxation order 2 was used.

Remark 6.1. *The polynomial optimization problem (P_0) has a close relationship with the problem of finding the solution with the least number of nonzero components which satisfies a system of polynomial inequalities and simple bounds. Mathematically, the problem of finding the solution with the least number of nonzero components which satisfies a system* *of polynomial inequalities and simple bounds, can be formulated as*

$$
\begin{array}{lll}
(P_0') & \min_{(x^1,\ldots,x^q)\in\mathbb{R}^{n_1}\times\ldots\times\mathbb{R}^{n_q}} & \sum_{l=1}^{q} \|x^l\|_0 \\
& s.t. & g_j^l(x^l) \le 0, j = 1,\ldots,m,\ l = 1,\ldots,q, \\
& & \|x^l\|_2^2 \le M^l,\ \ l = 1,\ldots,q,
\end{array}
$$

where $n_l \in \mathbb{N}$, $l = 1,\ldots,q$, and $\|x\|_0$ denotes the l_0-seminorm of the vector $x \in \mathbb{R}^n$, which gives the number of nonzero components of the vector x.

In the case where $q = 1$, $I_1 = \{1,\ldots,n\}$, $g_j(x) = a_j^T x - b_j$, $j = 1,\ldots,m$, $g_j(x) = -(a_j^T x - b_j)$, $j = m+1,\ldots,2m$, the problem (P_0') collapses to the Lasso problem with additional simple bounds which finds the solution with the least number of nonzero components satisfying simple bounds as well as linear equations $Ax = b$ with more unknowns than equalities:

$$
\begin{array}{lll}
(P_1) & \min_{x\in\mathbb{R}^n} & \|x\|_0 \\
& s.t. & Ax = b, \\
& & \|x\|_2^2 \le M,
\end{array}
$$

where $A = (a_1,\ldots,a_m)^T \in \mathbb{R}^{m\times n}$ $(m \le n)$, $b = (b_1,\ldots,b_m)^T \in \mathbb{R}^m$. We note that the standard Lasso problem which is given by

$$
\begin{array}{lll}
(P_2) & \min_{x\in\mathbb{R}^n} & \|x\|_0 \\
& s.t. & Ax = b,
\end{array}
$$

arises in signal processing and was examined, for example, in Candés and Tao (2005) [1]. Moreover, problem (P_1) and (P_2) has the same optimal value if $M > \|x^*\|_2^2$ for some solution x^* of problem (P_2).

In fact, the problem (P_0) and problem (P_0') are equivalent in the sense that $\min(P_0) = \min(P_0')$ *and $(x^{1*},\ldots,x^{q*}) \in \mathbb{R}^{n_1}\times\ldots\times\mathbb{R}^{n_q}$ is a solution of problem (P_0) if and only if $(x^{l*}, w^{l*}) \in \mathbb{R}^{n_l}\times\mathbb{R}^{n_l}$, $l = 1,\ldots,q$, is a solution of problem (P_0') where $w^{l*} = (w_{i_1}^{l*},\ldots,w_{i_{n_l}}^{l*}) \in \mathbb{R}^{n_l}$ is defined by*

$$
w_i^{l*} = \begin{cases} 1 & \text{if} \quad x_i^{l*} \neq 0, \\ 0 & \text{if} \quad x_i^{l*} = 0. \end{cases} \qquad i \in I_l,\ \ l = 1,\ldots,q. \tag{6.5}
$$

To see this, note that, for any solution $(x^{1*},\ldots,x^{q*}) \in \mathbb{R}^{n_1}\times\ldots\times\mathbb{R}^{n_q}$ of problem (P_0), let $w^{l*} \in \mathbb{R}^{n_l}$, $l = 1,\ldots,q$, be defined as in (6.5). Then, for each $l = 1,\ldots,q$, $(1-w_i^{l*})x_i^{l*} = 0$ and

$$
\sum_{i\in I_l} (w_i^{l*})^2 = \|x^{l*}\|_0.
$$

So, $\min(P_0) \ge \min(P_0')$. *Conversely, let (x^{l*}, w^{l*}), $l = 1,\ldots,q$, be a solution of problem (P_0). Then, x^{l*}, $l = 1,\ldots,q$, is feasible for problem (P_0'), and for all (x^l, w^l) feasible for problem (P_0), $\|x^{l*}\|_0 = \sum_{i=1}^n (w_i^{l*})^2 \le \sum_{i=1}^n (w_i^l)^2$. Now take any feasible point x^l,*

$l = 1,\dots,q$ for problem (P_0). Note that $(x^l, w^l(x^l))$, $l = 1,\dots,q$, is feasible for (P_0'), where $w^l(x^l)$ is given by

$$w^l(x^l)_i = \begin{cases} 1 & \text{if} \quad x_i^l \neq 0, \\ 0 & \text{if} \quad x_i^l = 0, \end{cases} \qquad i \in I_l,$$

and, for each $l = 1,\dots,q$,

$$\sum_{i\in I_l} (w^l(x^l)_i)^2 = \|x^l\|_0.$$

This implies that $\|x^{l*}\|_0 \leq \|x^l\|_0$ for any feasible point x^l, $l = 1,\dots,q$, of problem (P_0), and hence x^{l*} is a solution for problem (P_0). This yields $\min(P_0) \leq \min(P_0')$, thus, $\min(P_0) = \min(P_0')$. The remaining assertion follows from the construction of w^{l*}.

References

[1] E.J. Candés and T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory, 51 4203-4215.

[2] E. J. Candés, M. B. Wakin, and S. Boyd, Enhancing sparsity by reweighted l_1 minimization, J. Four. Anal. Appl., 14(5), 877-905

[3] M. Fukuda, M. Kojima, K. Murota, K. Nakata, *Exploiting sparsity in semidefinite programming via matrix completion I: General framework*, SIAM J. Optim., 11 (2000) 647-674.

[4] H. V. Hà, and T. S. Phạm, *Representations of positive polynomials and optimization on noncompact semi-algebraic sets,* SIAM J. Optim., 20 (2010), 3082-3103.

[5] V. Jeyakumar, J.B. Lasserre and G. Li, *On polynomial optimization over non-compact semi-algebraic sets*, J. Optim. Theo. Appl., (2014), DOI: 10.1007/s10957-014-0545-3 (available also on arXiv:1304.4552, July 2013)

[6] V. Jeyakumar, G. Li, and T.S. Pham, *Convergence of the Lasserre hierarchy of SDP relaxations for convex polynomial programs without compactness*, Oper. Res. Let., 42 (2014), 34-40.

[7] S. Kim, M. Kojima, M. Mevissen, and M. Yamashita, *Exploiting sparsity in linear and nonlinear matrix inequalities via positive semidefinite matrix completion*, Math. Program., 129 (2011) 33–68.

[8] S. Kim, M. Kojima, H. Waki, and M. Yamashita, *Algorithm 920: SFSDP: a Sparse Version of Full SemiDefinite Programming relaxation for sensor network localization problems*, ACM Tran. Math. Soft., 38(2012) 1-19.

[9] M. Kojima and M. Muramatsu, *A note on sparse SOS and SDP relaxations for polynomial optimization problems over symmetric cones*, Comp. Optim. Appl. 42(2009), no. 1, 31-41.

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

[11] J. B. Lasserre, *Moments, positive polynomials and their applications*, Imperial College Press, 2009.

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

[13] Z.-Q. Luo, N. D. Sidiropoulos, P. Tseng, and S. Zhang *Approximation Bounds for Quadratic Optimization with Homogeneous Quadratic Constraints* SIAM J. Optim., 18 (2007) 1-28.

[14] M. Marshall, *Positive polynomials and sums of squares,* Mathematical Surveys and Monographs 146. Providence, RI: American Mathematical Society, (2008).

[15] J.W. Nie, *An exact Jacobian SDP relaxation for polynomial optimization* Math Program, Ser. A, 137 (2013), no. 1-2, 225-255.

[16] J. Nie, *Optimality conditions and finite convergence of Lasserre's hierarchy*, Math Program, DOI: DOI 10.1007/s10107-013-0680-x, (2013).

[17] P. A. Parrilo, *Structured semidefinite programs and semi-algebraic geometry methods in robustness and optimization,* Ph.D. thesis, California Institute of Technology, May 2000.

[18] M. Putinar, *Positive polynomials on compact semi-algebraic sets*, Ind. Uni. Math. J. 41 (1993), 49-95.

[19] C. Scheiderer *Sums of squares on real algebraic curves.* Math. Zeit., 245 (2003), 725-760.

[20] K. Schmüdgen, *The K-moment problem for compact semi-algebraic sets.* Math. Ann., 289 (1991), 203-206.

[21] M. Schweighofer, *Global optimization of polynomials using gradient tentacles and sums of squares*, SIAM J. Optim. 17 (2006), 920-942.

[22] J. F. Strum, *SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones,* Optim. Methods Softw., 11 & 12(1999) 625–653 .

[23] H. Waki, S. Kim, M. Kojima and M. Muramatsu, *Sums of squares and semidefinite programming relaxations for polynomial optimization problems with structured sparsity*, SIAM J. Optim., 17 (2006), 218-242.

[24] H. Waki, S. Kim, M. Kojima, M. Muramtasu, H. Sugimoto, *Algorithm 883: SparsePOP: A sparse semidefinite programming relaxation of polynomial pptimization problems*, ACM Trans. Math. Softw, 35 (2008) 15.