

Convexification of polynomial optimization problems by means of monomial patterns

Gennadiy Averkov · Benjamin Peters · Sebastian Sager

Received: date / Accepted: date

Abstract Convexification is a core technique in global polynomial optimization. Currently, two different approaches compete in practice and in the literature. First, general approaches rooted in nonlinear programming. They are comparatively cheap from a computational point of view, but typically do not provide good (tight) relaxations with respect to bounds for the original problem. Second, approaches based on sum-of-squares and moment relaxations. They are typically computationally expensive, but do provide tight relaxations. In this paper, we embed both kinds of approaches into a unified framework of monomial relaxations. We develop a convexification strategy that allows to trade off the quality of the bounds against computational expenses. Computational experiments show that a combination with a prototype cutting-plane algorithm gives very encouraging results.

Keywords convexification · cutting-planes · McCormick relaxation · moment problem · nonlinear optimization · polynomial optimization · separation problem · sum-of-squares

1 Introduction

Many important convexification techniques applied to polynomial optimization problems share the following common distinctive features: in the case of a problem in n variables $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$, monomials

$$\mathbf{x}^\alpha := \mathbf{x}_1^{\alpha_1} \cdot \dots \cdot \mathbf{x}_n^{\alpha_n}$$

with $\alpha \in \mathbb{N}^n$ are substituted with *monomial variables* \mathbf{v}_α and the relationship between different monomial variables is captured, exactly or in a relaxed

G. Averkov, B. Peters, S. Sager
Faculty of Mathematics
Otto-von-Guericke University Magdeburg
E-mail: {gennadiy.averkov, benjamin.peters, sager}@ovgu.de

fashion, by systems of convex constraints. In order to describe the relationship between different monomial variables by constraints one needs to introduce additional *auxiliary monomial variables*.

Different approaches exist on how to pick these auxiliary monomial variables and the respective convex constraints. The nonlinear optimization community uses monomial variables and constraints such that the resulting relaxations are rather cheap to compute. Examples are McCormick relaxations [23,9], polyhedral outer approximations [29] or the α BB method [4,2,1]. The resulting poor lower bounds are compensated by calculating many relaxations within a branch-and-bound framework. The polynomial optimization community usually aims to solve only one single relaxation, which however produces a very tight bound. This often comes at the price of a large number of monomial variables and hard constraints. Examples are moment relaxation and sum-of-squares relaxation [5,20,21]. We propose a flexible template for the relaxation of polynomial problems allowing to trade off between the quality and computational costs of relaxations. We consider groups of monomial variables $\mathbf{v}_{\alpha^1}, \dots, \mathbf{v}_{\alpha^l}$ based on *patterns* $P = \{\alpha^1, \dots, \alpha^l\}$ of monomial exponents.

Example 1 For illustration, we consider a polynomial $f^{\text{ex1}} : \mathbb{R}^2 \mapsto \mathbb{R}$ defined by $f^{\text{ex1}} = \mathbf{f}_{0,2}\mathbf{x}_2^2 + \mathbf{f}_{1,1}\mathbf{x}_1\mathbf{x}_2 + \mathbf{f}_{2,3}\mathbf{x}_1^2\mathbf{x}_2^3 + \mathbf{f}_{2,4}\mathbf{x}_1^2\mathbf{x}_2^4 + \mathbf{f}_{4,0}\mathbf{x}_1^4 + \mathbf{f}_{5,5}\mathbf{x}_1^5\mathbf{x}_2^5$. The set $A = \text{supp}(\mathbf{f}) = \{(0,2), (1,1), (2,3), (2,4), (4,0), (5,5)\}$ corresponding to the coefficient vector $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^2}$ is important for different relaxation techniques. Patterns of different relaxation strategies for f^{ex1} are visualized in Figure 1.

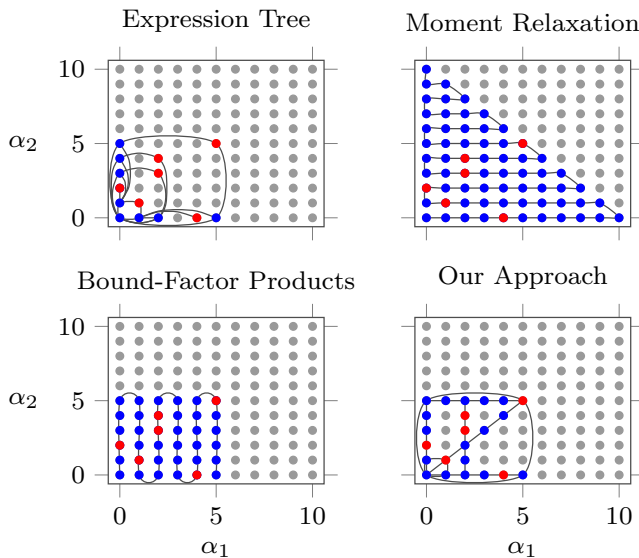


Fig. 1 Exponents $\alpha \in A \subseteq \mathbb{N}^2$ such that \mathbf{x}^α occurs in f^{ex1} (red points) and auxiliary moment variables (blue) for different exemplary relaxations. A pattern P is depicted as an undirected smooth curve passing through all the points of P .

The convex relaxation of the underlying problem is built by choosing an appropriate family of patterns $\cup_i P_i \supseteq A$ and linking the monomial variables within each pattern. For a detailed discussion of symbolic reformulation using expression trees, RLT using bound-factor products, moment relaxation, and our pattern relaxation from this new point of view see Section 5.

The paper is organized as follows. After explaining the basic notation in Section 2, we introduce in Section 3 the notion of the pattern relaxations and formulate the separation problem for patterns as an optimization problem. In Section 4 we introduce different pattern types. Coming back to the examples from Figure 1, we formulate established convexification techniques from the new point of view in Section 5. In Section 6 we present two novel algorithms that compute patterns and lower bounds for (POP) using pattern relaxations and cutting-planes for the important case of minimizing a polynomial function f over a box. We discuss the obtained computational results in Section 7. Finally, a conclusion is given in Section 8.

2 Basic Notation

As usual, the set of natural numbers including zero is denoted by \mathbb{N} , for a positive integer n the set $\{1, \dots, n\}$ by $[n]$ and the set $[n] \cup \{\mathbf{0}\}$ by $[n]_0$. We define the bilinear product of two vectors $\mathbf{v} \in \mathbb{R}^A$ and $\mathbf{w} \in \mathbb{R}^B$ with $A, B \subseteq \mathbb{N}^n$ as

$$\langle \mathbf{v}, \mathbf{w} \rangle := \sum_{\alpha \in A \cap B} \mathbf{v}_\alpha \mathbf{w}_\alpha.$$

The standard basis vectors of \mathbb{R}^A are denoted by \mathbf{e}^α for $\alpha \in A$. The coordinate projection of $\mathbf{v} \in \mathbb{R}^A$ onto components indexed by a nonempty subset P of A is

$$\mathbf{v}_P := (\mathbf{v}_\alpha)_{\alpha \in P}.$$

The l_1 norm of \mathbf{v} is $\|\mathbf{v}\|_1$ and the l_∞ norm of \mathbf{v} is $\|\mathbf{v}\|_\infty$. For a nonempty and compact set $X \subseteq \mathbb{R}^A$, vectors $\mathbf{v}, \mathbf{c} \in \mathbb{R}^A$ and $\varepsilon > 0$ we call

$$N_\varepsilon(X) := \{\mathbf{v} \in \mathbb{R}^A : \|\mathbf{v} - \mathbf{u}\|_1 \leq \varepsilon \text{ for some } \mathbf{u} \in X\}$$

the ε -neighbourhood of X ,

$$\text{diam}(X) := \max_{\mathbf{v}, \mathbf{u} \in X} \|\mathbf{v} - \mathbf{u}\|_1,$$

the *diameter* of X ,

$$\omega_X(\mathbf{c}) := \max\{\langle \mathbf{c}, \mathbf{v} \rangle : \mathbf{v} \in X\} - \min\{\langle \mathbf{c}, \mathbf{v} \rangle : \mathbf{v} \in X\}$$

the *width function* of X in direction \mathbf{c} and

$$\text{dist}(X, \mathbf{v}) := \min_{\mathbf{u} \in X} \|\mathbf{v} - \mathbf{u}\|_1$$

the *distance function* of X and \mathbf{v} . From the triangle inequality it follows that the distance function is continuous in \mathbf{v} . Under a *covering* of an interval $[a, b]$ we understand a finite family of segments \mathcal{I} satisfying

$$[a, b] = \bigcup_{I \in \mathcal{I}} I.$$

The *fineness* of \mathcal{I} is defined as

$$\varrho(\mathcal{I}) := \max \{u - l : [l, u] \in \mathcal{I}\}.$$

We define the *support* of a vector $\alpha \in \mathbb{N}^n$ and the *support* of a set $P \subseteq \mathbb{N}^n$

$$\text{supp}(\alpha) := \{i \in [n] : \alpha_i \neq 0\} \quad \text{and} \quad \text{supp}(P) := \bigcup_{\alpha \in P} \text{supp}(\alpha).$$

A vector α is said to have *full support* if $\text{supp}(\alpha) = [n]$. The *degree* of the set P is

$$\text{deg}(P) := \max\{\|\alpha\|_1 : \alpha \in P\}.$$

The minimum and maximum of the monomial \mathbf{x}^α over the box K are

$$\mathbf{x}_{\min}^\alpha := \min_{\mathbf{x} \in K} \mathbf{x}^\alpha \quad \text{and} \quad \mathbf{x}_{\max}^\alpha := \max_{\mathbf{x} \in K} \mathbf{x}^\alpha.$$

The *moment vector map* of a set $A \subseteq \mathbb{N}^n$ is

$$m^A(\mathbf{x}) := (\mathbf{x}^\alpha)_{\alpha \in A}.$$

A polynomial $p \in \mathbb{R}[\mathbf{x}]$ is called *sum-of-squares* (SOS), if $p = p_1^2 + \dots + p_k^2$ for finitely many polynomials $p_1, \dots, p_k \in \mathbb{R}[\mathbf{x}]$. As usual, we use $\Sigma_{n,2d}$ to denote the cone of n -variate SOS polynomials of degree at most $2d$ and *psd* abbreviates positive semidefinite.

3 Pattern Relaxation

3.1 Monomial Convexification and Monomial Relaxation

We consider polynomials f whose coefficient vector $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n}$ satisfies $\text{supp}(\mathbf{f}) \subseteq A$, where A is a given set. That is, by A we prescribe which monomials can occur in f . The feasible set of our polynomial problem is a box

$$K := [\mathbf{a}_1, \mathbf{b}_1] \times \dots \times [\mathbf{a}_n, \mathbf{b}_n] \subseteq \mathbb{R}^n$$

with $\mathbf{a}_i < \mathbf{b}_i$. We can state our problem as

$$\begin{aligned} & \text{minimize} && f(\mathbf{x}) \\ & \text{for} && \mathbf{x} \in \mathbb{R}^n \\ & \text{subject to} && \mathbf{x} \in K. \end{aligned} \tag{POP}$$

Via lifting, we reformulate (POP) as a problem of minimizing a linear functional on \mathbb{R}^A :

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^A \\ & \text{subject to} && \mathbf{v} \in \{m^A(\mathbf{x}) : \mathbf{x} \in K\}. \end{aligned}$$

Replacing the feasible set by its convex hull

$$M^A(K) := \text{conv}(\{m^A(\mathbf{x}) : \mathbf{x} \in K\})$$

yields the *monomial convexification* of (POP):

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^A \\ & \text{subject to} && \mathbf{v} \in M^A(K) \end{aligned} \tag{C-POP}$$

We refer to $M^A(K)$ as (*n-variate*) *moment body*. Clearly, the convexification (C-POP) of (POP) is tight, that is, the optimal values of (C-POP) and (POP) coincide. For general sets A , the constraint $\mathbf{v} \in M^A(K)$ is difficult to deal with. Thus, it is natural to relax $\mathbf{v} \in M^A(K)$ to a system of simpler constraints of the same type

$$\mathbf{v}_{P_i} \in M^{P_i}(K) \text{ for } i \in [m], \tag{1}$$

where the sets P_i satisfy

$$A \subseteq P_1 \cup \dots \cup P_m. \tag{2}$$

We call P_i a *pattern* and (1) the *pattern relaxation* of $M^A(K)$ with respect to the family of patterns $\{P_1, \dots, P_m\}$. Throughout the paper we use \bar{A} to denote $P_1 \cup \dots \cup P_m$. Using the pattern relaxation of $M^A(K)$ we obtain a lower bound on (POP) by solving

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\bar{A}} \\ & \text{subject to} && \mathbf{v}_{P_i} \in M^{P_i}(K) \text{ for all } i \in [m]. \end{aligned} \tag{P-RLX}$$

Note that (P-RLX) is bounded whenever $\mathbf{f} \in \mathbb{R}^{\mathbb{N}^n}$ satisfies $\text{supp}(\mathbf{f}) \subseteq \bar{A}$.

The advantage of the above approach is that we can decide how to choose patterns P_1, \dots, P_m in a way that we can meet our requirements on the computational costs needed to solve the respective instance of (P-RLX). We solve (P-RLX) using a cutting-plane algorithm that iteratively generates cuts for the sets $M^{P_i}(K)$. The computational costs of the cutting-plane algorithm are thus directly related to the costs of solving separation problems for $M^{P_i}(K)$. In view of this, we are primarily interested in the choice of patterns P_i , for which the computational costs of generating a cut for $M^{P_i}(K)$ meet our requirements. Since (2) is an inclusion and not an equality, we can find such

patterns even if A is ill-structured. This explains the reason for introducing additional variables in \mathbf{v}_β , $\beta \in \bar{A} \setminus A$ in (P-RLX), which have not been present in (C-POP).

The entire procedure can also be viewed as embedding $M^A(K)$ into $M^{\bar{A}}(K)$ for some set \bar{A} that contains A and can be represented nicely as a union of patterns P_1, \dots, P_m . Phrased geometrically, the passage from (POP) through (C-POP) to (P-RLX) can be represented by the diagram

$$m^A(K) \xrightarrow{\text{convexifying}} M^A(K) \xrightarrow{\text{embedding}} M^{\bar{A}}(K) \xrightarrow{\text{projecting}} M^{P_i}(K).$$

The quality of a pattern relaxation of $M^A(K)$ with respect to the family of patterns $\mathcal{P} := \{P_1, \dots, P_m\}$ depends on how the moment variables are connected by the system of conditions (1). We say that monomial variables $\mathbf{v}_\alpha, \mathbf{v}_\beta$ are *directly connected* by \mathcal{P} if $\alpha, \beta \in P_i \setminus \{\mathbf{0}\}$ holds for some $i \in [m]$. Furthermore, $\mathbf{v}_\alpha, \mathbf{v}_\beta$ are *indirectly connected* by \mathcal{P} if, for some finitely many indices $i_1, \dots, i_k \in [m]$, one has $\alpha \in P_{i_1} \setminus \{\mathbf{0}\}, \beta \in P_{i_k} \setminus \{\mathbf{0}\}$ and $P_{i_j} \cap P_{i_{j+1}} \setminus \{\mathbf{0}\} \neq \emptyset$ for all $j \in [k-1]$.

3.2 Separation Problem

We use a cutting-plane algorithm to solve (P-RLX) that generates valid inequalities for $M^P(K)$ from the following maximization problem.

$$\begin{aligned} & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{c} \in \mathbb{R}^P \text{ and } \delta \in \mathbb{R} \\ & \text{subject to} && \langle \mathbf{c}, m^P(\mathbf{x}) \rangle \geq \delta \quad \text{for all } \mathbf{x} \in K, \\ & && \mathbf{c}_\alpha \in [-1, 1] \quad \text{for all } \alpha \in P. \end{aligned} \tag{SP}$$

If \mathbf{v} is not in $M^P(K)$, then (SP) has a positive optimal value and the optimal solution \mathbf{c}^*, δ^* of (SP) yields the inequality $\langle \mathbf{c}^*, \mathbf{u} \rangle \geq \delta^*$, which is valid for all $\mathbf{u} \in M^P(K)$ and violated for $\mathbf{u} = \mathbf{v}$. Note that the equality $\langle \mathbf{c}^*, \mathbf{u} \rangle = \delta^*$ defines a supporting hyperplane of $M^P(K)$ that contains a point of $M^P(K)$ closest to \mathbf{v} in the l_1 -norm:

Proposition 1 ([10, Ch. 8.1.3]) *Let P be a pattern, $\mathbf{v} \in \mathbb{R}^P$. Then the optimal value of (SP) is $\text{dist}(M^P(K), \mathbf{v})$.*

4 Pattern Types

In order to generate computational tractable pattern relaxations of (POP) we need to find patterns P such that we can formulate the constraint

$$\langle \mathbf{c}, m^P(\mathbf{x}) \rangle \geq \delta \text{ for all } \mathbf{x} \in K$$

of (SP) in such a way that it is accessible to optimization methods. In this section we introduce four useful types of patterns for which we can handle

this constraint. In what follows, the sets A_1, \dots, A_6 from Figure 2 are used to illustrate the relaxations of the moment bodies $M^{A_i}(K)$ based on our pattern types.

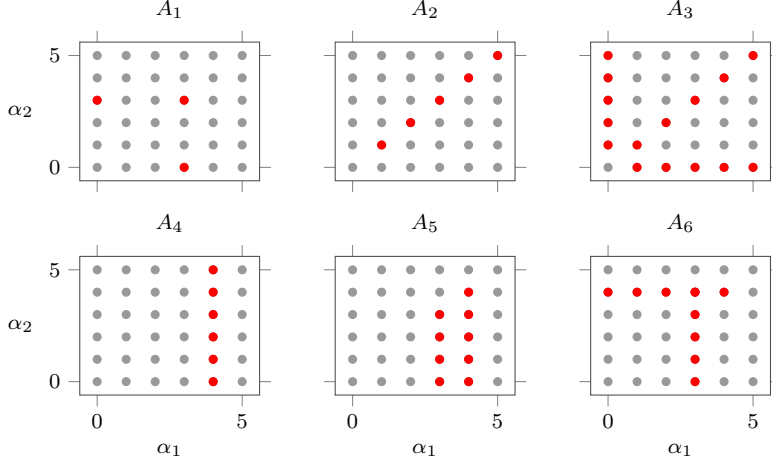


Fig. 2 The sets A_1, \dots, A_6 , depicted in red, used to illustrate different pattern types.

4.1 Singleton Pattern

The smallest patterns are *singeltons* $\{\alpha\}$ with $\alpha \in \mathbb{N}^n$. The moment body of the singleton is the interval $M^{\{\alpha\}}(K) = [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha]$. The pattern relaxation of $M^A(K)$ induced by the family of singletons $\{P_1, \dots, P_m\} = \{\{\alpha\} : \alpha \in A\}$ is the box $\{\mathbf{v} \in \mathbb{R}^A : \mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha]\}$. We can solve (P-RLX) exactly in this case, as the optimum is attained at the vertex $\mathbf{v} \in \mathbb{R}^A$ with $\mathbf{v}_\alpha = \mathbf{x}_{\min}^\alpha$ if $\mathbf{f}_\alpha \geq 0$ and $\mathbf{v}_\alpha = \mathbf{x}_{\max}^\alpha$ if $\mathbf{f}_\alpha < 0$. This is the weakest possible relaxation within the pattern approach.

4.2 Multilinear Pattern

We introduce the *Hadamard product* of vectors $\alpha, \omega \in \mathbb{N}^n$ by $\alpha \circ \omega := (\alpha_i \omega_i)_{i \in [n]}$. We call

$$\text{ML}(\alpha) := \{\alpha \circ \omega \in \mathbb{N}^n : \omega \in \{0, 1\}^n\},$$

with $\alpha \in \mathbb{N}^n$, a *multilinear pattern* (ML). See Figure 3 for an illustration.

Proposition 2 *Let $\alpha \in \mathbb{N}^n$ be of full support. The moment body $M^{\text{ML}(\alpha)}(K)$ is a polytope satisfying*

$$M^{\text{ML}(\alpha)}(K) = \text{conv}(m^{\{0,1\}^n}(V))$$

with $V := \{\mathbf{x}_{\min}^{\alpha_1 \mathbf{e}^1}, \mathbf{x}_{\max}^{\alpha_1 \mathbf{e}^1}\} \times \cdots \times \{\mathbf{x}_{\min}^{\alpha_n \mathbf{e}^n}, \mathbf{x}_{\max}^{\alpha_n \mathbf{e}^n}\}$.

Proof Using $\tilde{K} := [\mathbf{x}_{\min}^{\alpha_1 \mathbf{e}^1}, \mathbf{x}_{\max}^{\alpha_1 \mathbf{e}^1}] \times \cdots \times [\mathbf{x}_{\min}^{\alpha_n \mathbf{e}^n}, \mathbf{x}_{\max}^{\alpha_n \mathbf{e}^n}]$ we can write

$$m^{\text{ML}(\alpha)}(K) = m^{\{0,1\}^n}(\tilde{K}).$$

Clearly, $m^{\{0,1\}^n}(\mathbf{x})$ is a multilinear map. So, if \mathbf{x}^* is in \tilde{K} , but not in the vertex set V of \tilde{K} , there exists $i \in [n]$ such that the points $\mathbf{x}^* \pm \varepsilon \mathbf{e}^i$ belong to \tilde{K} , for a sufficiently small $\varepsilon > 0$. By multilinearity of $m^{\{0,1\}^n}(\mathbf{x})$ one has

$$m^{\{0,1\}^n}(\mathbf{x}^*) = \frac{1}{2} m^{\{0,1\}^n}(\mathbf{x}^* - \varepsilon \mathbf{e}^i) + \frac{1}{2} m^{\{0,1\}^n}(\mathbf{x}^* + \varepsilon \mathbf{e}^i)$$

This shows that the points of $\tilde{K} \setminus V$ are not extreme points of $m^{\{0,1\}^n}(\tilde{K})$. Consequently, $m^{\{0,1\}^n}(\tilde{K}) = m^{\{0,1\}^n}(V)$. \square

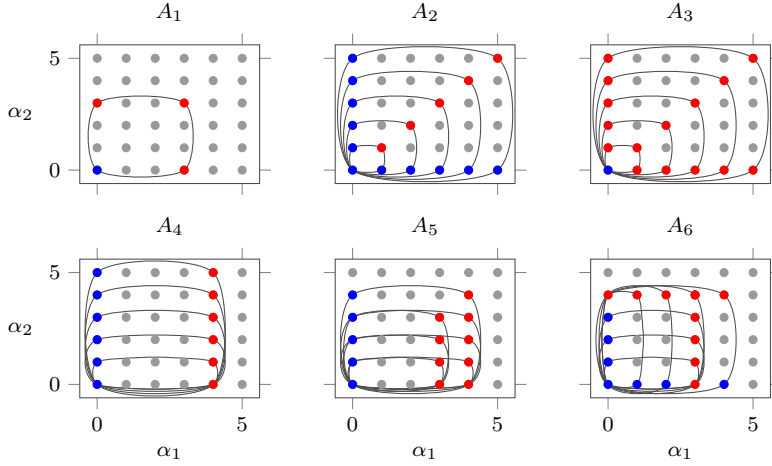


Fig. 3 Multilinear patterns applied to A_1, \dots, A_6 ; formatting as in Figure 1.

Corollary 1 *In the setting of Proposition 2, the separation problem for the moment body $M^{\text{ML}(\alpha)}(K)$ and a point $\mathbf{v} \in \mathbb{R}^{\text{ML}(\alpha)}$ can be formulated as the linear program*

$$\begin{aligned} & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{c} \in \mathbb{R}^{\text{ML}(\alpha)} \text{ and } \delta \in \mathbb{R} \\ & \text{subject to} && \sum_{\omega \in \{0,1\}^n} \mathbf{c}_{\alpha \circ \omega} \mathbf{w}_\omega \geq \delta \quad \text{for all } \mathbf{w} \in m^{\{0,1\}^n}(V), \quad (3) \\ & && \mathbf{c}_\beta \in [-1, 1] \quad \text{for all } \beta \in \text{ML}(\alpha). \end{aligned}$$

Proof The assertion follows from Proposition 2. \square

The problem (3) involves 2^n inequalities indexed by vectors $\mathbf{w} \in m^{\{0,1\}^n}(V)$. If α is not of full support, the respective separation problem (SP) can be formulated analogously with $2^{|\text{supp}(\alpha)|}$ inequalities.

4.3 Truncated Submonoid Pattern

Let $B \subseteq \mathbb{N}^n$ be a nonempty finite set and $\Gamma = (\gamma^1, \dots, \gamma^k) \in \mathbb{N}^{n \times k}$, $k \in [n]$, be a matrix, whose columns $\gamma^i \in B$ are nonzero vectors with pairwise disjoint supports. Clearly, such vectors $\gamma^1, \dots, \gamma^k$ are linearly independent. We denote by $\Lambda^+(\Gamma)$ the submonoid of $(\mathbb{N}^n, +, \mathbf{0})$ generated by the columns of Γ , that is,

$$\Lambda^+(\Gamma) := \gamma^1 \mathbb{N} + \dots + \gamma^k \mathbb{N}.$$

We call

$$\text{TS}(\Gamma, B) := \Lambda^+(\Gamma) \cap B,$$

the k -variate B -truncated submonoid pattern (TS). In computations, we use $\text{TS}(\Gamma, B)$ with B being the discrete box

$$B = \{0, \dots, \beta_1\} \times \dots \times \{0, \dots, \beta_n\}, \quad (4)$$

where $\beta \in \mathbb{N}^n$.

Proposition 3 *Let $B \subseteq \mathbb{N}^n$ be finite and nonempty set and $\Gamma \in \mathbb{N}^{n \times k}$ be a matrix with columns $\gamma^1, \dots, \gamma^k \in B \setminus \{\mathbf{0}\}$ having pairwise disjoint supports. Then the moment body $M^{\text{TS}(\Gamma, B)}(K)$ can be represented as a k -variate moment body by*

$$M^{\text{TS}(\Gamma, B)}(K) = M^{\tilde{P}}(\tilde{K}),$$

with

$$\tilde{P} = \{\omega \in \mathbb{N}^k : \Gamma \omega \in \text{TS}(\Gamma, B)\} \subseteq \mathbb{N}^k$$

and

$$\tilde{K} = [\mathbf{x}_{\min}^{\gamma^1}, \mathbf{x}_{\max}^{\gamma^1}] \times \dots \times [\mathbf{x}_{\min}^{\gamma^k}, \mathbf{x}_{\max}^{\gamma^k}].$$

Proof The desired representation is obtained by taking the convex hull of the left and the right hand side of the equality $m^{\text{TS}(\Gamma, B)}(K) = m^{\tilde{P}}(\tilde{K})$. \square

Corollary 2 *In the setting of Proposition 3, the separation problem (SP) for the moment body $M^{\text{TS}(\Gamma, B)}(K)$ and a point $\mathbf{v} \in \mathbb{R}^{\text{TS}(\Gamma, B)}$ can be formulated as follows:*

$$\begin{aligned} & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{c} \in \mathbb{R}^{\text{TS}(\Gamma, B)} \text{ and } \delta \in \mathbb{R} \\ & \text{subject to} && p_{\mathbf{c}}(\tilde{\mathbf{x}}) \geq \delta \quad \text{for all } \tilde{\mathbf{x}} \in \tilde{K}, \\ & && \mathbf{c}_{\beta} \in [-1, 1] \quad \text{for all } \beta \in \text{TS}(\Gamma, B), \end{aligned} \quad (5)$$

where $p_{\mathbf{c}}(\tilde{\mathbf{x}})$ is a k -variate polynomial given by

$$p_{\mathbf{c}}(\tilde{\mathbf{x}}) = \sum_{\omega \in \tilde{P}} \mathbf{c}_{\Gamma \omega} \tilde{\mathbf{x}}^{\omega}.$$

Proof The assertion follows from Proposition 3 and $M^{\tilde{P}}(\tilde{K}) = \text{conv}(m^{\tilde{P}}(\tilde{K}))$. \square

Naturally, computability of (5) depends on the number of the variables k and the degree of the polynomial $p_c(\tilde{\mathbf{x}})$. In other words, computability of (5) depends on the degree of the set \tilde{P} and the number of submonoid generators $\gamma^1, \dots, \gamma^k$. For practical purposes, it is desirable to choose k to be a relatively small number. Furthermore, by choosing $\gamma^1, \dots, \gamma^k \in B \setminus \{\mathbf{0}\}$ to be long vectors we can keep the degree of \tilde{P} small. All this allows to control the tractability of (5). We would like to stress that in practice it is usually infeasible to use SOS

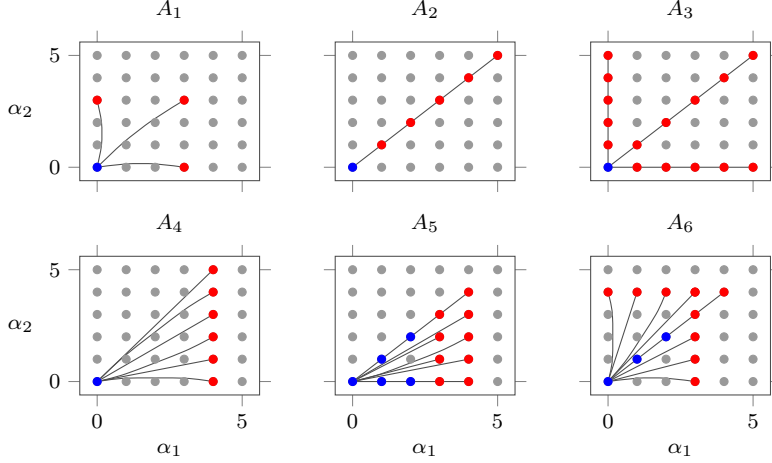


Fig. 4 Chain patterns applied to A_1, \dots, A_6 ; formatting as in Figure 1.

relaxations for the original problem (POP) due to the size of these relaxations; for a theoretical justification see also [7]. In contrast, we believe that one can use SOS relaxations for (5), since we can keep the size of (5) under control. Applying the following results from real algebraic geometry, we can build a SOS relaxation (6) of (5).

Theorem 1 *Let $p(\mathbf{x})$ be a polynomial satisfying $p(\mathbf{x}) > 0$ for all $\mathbf{x} \in K$. Then there exists an even integer d and polynomials $\sigma_0 \in \Sigma_{n,d}$ and $\sigma_1, \dots, \sigma_n \in \Sigma_{n,d-2}$ such that*

$$p = \sigma_0 + \sum_{i=1}^n \sigma_i(\mathbf{b}_i - \mathbf{x}_i)(\mathbf{x}_i - \mathbf{a}_i).$$

Proof This is a special case of Putinar's theorem [24]. See [6] for a short proof.

Proposition 4 ([20, Ch. 2.1]) *The cone of n -variate sum-of-squares polynomials of degree at most $2d$ is the image of the cone of semidefinite matrices of size $k := \binom{n+d}{n}$ under a linear transformation. More precisely, one has*

$$\Sigma_{n,2d} = \{m^{\mathbb{N}_d^n}(\mathbf{x})^\top T m^{\mathbb{N}_d^n}(\mathbf{x}) : T \text{ is a positive semidefinite } k \times k \text{ matrix}\}.$$

Corollary 3 *In the setting of Corollary 2 let $g_i(\tilde{\mathbf{x}}) = (\mathbf{x}_{\max}^{\gamma_i} - \tilde{\mathbf{x}})(\tilde{\mathbf{x}} - \mathbf{x}_{\min}^{\gamma_i})$, $i \in [n]$. For every even integer $d \geq \deg(\text{TS}(\Gamma, B))$ consider the optimization problem*

$$\begin{aligned}
& \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
& \text{for} && \mathbf{c} \in \mathbb{R}^{\text{TS}(\Gamma, B)}, \delta \in \mathbb{R}, \\
& && \sigma_0 \in \Sigma_{k,d} \text{ and } \sigma_1, \dots, \sigma_n \in \Sigma_{k,d-2} \\
& \text{subject to} && p_{\mathbf{c}} - \delta = \sigma_0 + \sum_{i=1}^n \sigma_i g_i \\
& && \mathbf{c}_\beta \in [-1, 1] \quad \text{for all } \beta \in \text{TS}(\Gamma, B).
\end{aligned} \tag{6}$$

Then the following hold:

- (a) If $\mathbf{c}, \delta, \sigma_0, \dots, \sigma_n$ is a feasible solution of (6), then \mathbf{c}, δ is a feasible solution of (5).
- (b) As $d \rightarrow \infty$, the optimal solution of (6) converges to the optimal solution of (5).

Note that (6) is a conic optimization problem involving the SOS cones $\Sigma_{k,d}$ and $\Sigma_{k,d-2}$. By Proposition 4, the conic variables $\sigma_0, \dots, \sigma_n$ can be replaced by semidefinite matrix variables via lifting. This transformation turns (6) to a semidefinite problem.

4.4 Chain Pattern

For $\gamma \in \mathbb{N}^n \setminus \{\mathbf{0}\}$ and $d \in \mathbb{N}$, we call

$$\text{CH}(\gamma, d) = \{i\gamma : i = 0, \dots, d\}$$

a *chain pattern*. A chain pattern is a special truncated submonoid pattern with $k = 1$. In the case of chains, problem (5) amounts to the problem

$$\begin{aligned}
& \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
& \text{for} && \mathbf{c} \in \mathbb{R}^{\text{CH}(\gamma, d)} \text{ and } \delta \in \mathbb{R} \\
& \text{subject to} && p_{\mathbf{c}}(t) \geq \delta \quad \text{for all } t \in [\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma] \\
& && \mathbf{c}_\beta \in [-1, 1] \quad \text{for all } \beta \in \text{CH}(\gamma, d),
\end{aligned} \tag{7}$$

where $p_{\mathbf{c}}(t) = \sum_{i=0}^d \mathbf{c}_{i\gamma} t^i$. It is known that the nonnegativity of $p_{\mathbf{c}}(t) - \delta$ on $[\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma]$ can be reformulated as a semidefinite constraint:

Theorem 2 ([21, Th. 3.23]) *Let d be an even nonnegative integer, $a, b \in \mathbb{R}$ and $a < b$. Let p be a univariate polynomial of degree at most $2d$ with $p(t) \geq 0$ for all $t \in [a, b]$. Then $p = \sigma_0 + \sigma_1(t - a)(b - t)$ holds for some $\sigma_0 \in \Sigma_{1,d}$ and $\sigma_1 \in \Sigma_{1,d-2}$.*

As a direct consequence of Theorem 2 and Proposition 4, we obtain

Corollary 4 *Let $\gamma \in \mathbb{N}^n \setminus \{\mathbf{0}\}$. Let $g(t) = (\mathbf{x}_{\max}^\gamma - t)(t - \mathbf{x}_{\min}^\gamma)$ and d be an even integer. Then problem (7) can be formulated as*

$$\begin{aligned} & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{c} \in \mathbb{R}^{\text{CH}(\gamma, d)}, \delta \in \mathbb{R}, \\ & && \sigma_0 \in \Sigma_{1, d} \text{ and } \sigma_1 \in \Sigma_{1, d-2} \\ & \text{subject to} && p_{\mathbf{c}} - \delta = \sigma_0 + \sigma_1 g \\ & && \mathbf{c}_\beta \in [-1, 1] \quad \text{for all } \beta \in \text{CH}(\gamma, d). \end{aligned} \tag{8}$$

Another way to approach the separation problem for chains is by approximating the moment body $M^{\text{CH}(\gamma, d)}(K)$ by a polytope and solving a linear relaxation of (7). Proposition 5 shows how $M^{\text{CH}(\gamma, d)}(K)$ can be approximated by polytopes to arbitrary precision.

We will make use of the following notation. With each segment $[l, u] \subseteq \mathbb{R}$ we associate the $(d+1) \times (d+1)$ matrix

$$\Phi^{[l, u]} := \left(\binom{k}{j} l^{k-j} (u-l)^j \right)_{k, j \in [d]_0} \in \mathbb{R}^{[d]_0 \times [d]_0}$$

and the polytope

$$\Delta^{[l, u]} := \text{conv} \left\{ \Phi^{[l, u]}(\mathbf{u}^i) : i \in [d]_0 \right\},$$

where $\mathbf{u}^i := \sum_{k=0}^i \mathbf{e}^k$ for $i \in [d]_0$.

Proposition 5 *Let $d \in \mathbb{N}$. Let $[a, b] \subseteq \mathbb{R}$ be a segment and \mathcal{I} a covering of $[a, b]$. Then*

$$M^{[d]_0}([a, b]) \subseteq \text{conv} \left(\bigcup_{I \in \mathcal{I}} \Delta^I \right). \tag{9}$$

Furthermore, there exists a constant $C^* > 0$ depending only on d and the segment $[a, b]$ such that, for every $\varepsilon > 0$, the inequality $\varrho(\mathcal{I}) \leq C^* \varepsilon$ implies

$$\text{conv} \left(\bigcup_{I \in \mathcal{I}} \Delta^I \right) \subseteq N_\varepsilon(M^{[d]_0}([a, b])). \tag{10}$$

Before giving a proof of Proposition 5, we establish the following

Lemma 1 *Let $d \in \mathbb{N}$, let $l, u \in \mathbb{R}$ and $l < u$. Then one has*

$$\text{diam}(\Delta^{[l, u]}) \leq |u - l| d^2 \eta_{l, u}^d,$$

where

$$\eta_{l, u} := \max\{|l| + |u - l|, 1\}.$$

Proof Taking into account that the diameter of a set does not change by taking the convex hull, we arrive at the representation

$$\text{diam}(\Delta^{[l,u]}) = \max \left\{ \Phi^{[l,u]}(\mathbf{u}^j - \mathbf{u}^i) : i, j \in [d]_0, i < j \right\}. \quad (11)$$

Let $i, j \in [d]_0$ and $i < j$ and let $\mathbf{u} := \mathbf{u}^j - \mathbf{u}^i$. We derive an upper bound on $\|\mathbf{v}\|_1$ for $\mathbf{v} = \Phi^{[l,u]}\mathbf{u}$. One has

$$\mathbf{v}_k = \sum_{h=0}^k \binom{k}{h} l^{k-h} (u-l)^h \mathbf{u}_h \quad (12)$$

for $k \in [d]_0$. The vector \mathbf{u} belongs to $\{0, 1\}^{[d]_0}$ and its component \mathbf{u}_0 is 0. Since $\mathbf{u}_0 = 0$, we obtain that $\mathbf{v}_0 = 0$ and that the summand for $h = 0$ in the sum on the right hand side of (12) is zero. This yields

$$\begin{aligned} \|\mathbf{v}\|_1 &= |\mathbf{v}_1| + \dots + |\mathbf{v}_d| \\ &\leq \sum_{k=1}^d \sum_{h=1}^k \binom{k}{h} |l|^{k-h} |u-l|^h \\ &= \sum_{k=1}^d ((|l| + |u-l|)^k - |l|^k) \\ &= |u-l| \sum_{k=1}^d \sum_{h=0}^{k-1} (|l| + |u-l|)^h |l|^{k-1-h} \\ &\leq |u-l| \sum_{k=1}^d \sum_{h=0}^{k-1} \eta_{l,u}^{k-1} \\ &\leq |u-l| d^2 \eta_{l,u}^d. \end{aligned}$$

Applying this inequality to (11) yields the assertion. \square

We now prove Proposition 5.

Proof (Proposition 5) Consider an arbitrary segment $[l, u]$. We use the identity

$$\Phi^{[l,u]} m^{[d]_0}(t) = m^{[d]_0}((1-t)l + tu), \quad (13)$$

which holds for every $t \in \mathbb{R}$ and can be derived using the binomial expansion for the components of the right-hand side of (13). Recall that $\mathbf{u}^i = \sum_{k=0}^i \mathbf{e}^k$. The polytope $\text{conv}(\{\mathbf{u}^0, \dots, \mathbf{u}^d\})$ is a d -dimensional simplex, which can be described as the set of all $\mathbf{u} \in \mathbb{R}^{[d]_0}$ satisfying $1 = \mathbf{u}_0 \geq \mathbf{u}_1 \geq \dots \geq \mathbf{u}_d \geq 0$. Hence $m^{[d]_0}(t)$ belongs to $\text{conv}(\{\mathbf{u}^0, \dots, \mathbf{u}^d\})$ for every $t \in [0, 1]$. Consequently, $\Phi^{[l,u]} m^{[d]_0}(t)$ belongs to $\text{conv}(\{\Phi^{[l,u]}\mathbf{u}^0, \dots, \Phi^{[l,u]}\mathbf{u}^d\}) = \Delta^{[l,u]}$. Taking into account (13), we see that $m^{[d]_0}(t) \in \Delta^{[l,u]}$ holds for every $t \in [l, u]$. The latter immediately implies (9).

We now derive the second part of the assertion. In view of Lemma 1, if $[l, u] \in \mathcal{I}$, then every point of $\Delta^{[l, u]}$ has l_1 -distance at most $|u - l|d^2\eta_{l, u}^d$ to the point $m^{[d]_0}(l)$, which belongs to $M^{[d]_0}([a, b])$ and to $\Delta^{[l, u]}$. This yields

$$\Delta^{[l, u]} \subseteq N_\varepsilon(M^{[d]_0}([a, b]))$$

for every $\varepsilon \geq \varrho(\mathcal{I})d^2\eta_{l, u}^d$. Note that $\eta_{l, u} \leq \max\{|b| + |b - a|, 1\}$. We thus obtain

$$\bigcup_{I \in \mathcal{I}} \Delta^{[l, u]} \subseteq N_\varepsilon(M^{[d]_0}([a, b])) \quad (14)$$

for every $\varepsilon \geq \frac{\varrho(\mathcal{I})}{C^*}$ with $C^* = d^{-2}(\max\{|b| + |b - a|, 1\})^{-d}$. Since the right-hand side of the latter inclusion is a convex set, taking the convex hull of the left-hand side we see that the inclusion (10) holds when $\varepsilon > 0$ satisfies the inequality $\rho(\mathcal{I}) \geq C^*\varepsilon$. \square

Proposition 5 allows to solve the separation problem for $M^{\text{TS}(\gamma, B)}(K)$ approximately using linear programming:

Corollary 5 *Let $\gamma \in \mathbb{N}^n \setminus \{\mathbf{0}\}$ and let d be a positive integer. Then there exists a constant $C^* > 0$ that depends only on $\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma$ and d such that the following holds: If a vector $\mathbf{v} \in \mathbb{R}^{\text{CH}(\gamma, d)}$ does not belong to $M^{\text{CH}(\gamma, d)}(K)$ and \mathcal{I} is a covering of $[\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma]$ with*

$$\varrho(\mathcal{I}) < C^* \text{dist}(M^{\text{CH}(\gamma, d)}(K), \mathbf{v}),$$

then the optimal value of the linear program

$$\begin{aligned} & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{c} \in \mathbb{R}^{\text{CH}(\gamma, d)} \text{ and } \delta \in \mathbb{R} \\ & \text{subject to} && \sum_{i=0}^d \mathbf{c}_{i\gamma} \mathbf{w}_i \geq \delta && \text{for all } \mathbf{w} = \Phi^I \mathbf{u}^i \text{ with} && (15) \\ & && && I \in \mathcal{I} \text{ and } i \in [d]_0, \\ & && \mathbf{c}_\beta \in [-1, 1] && \text{for all } \beta \in \text{CH}(\gamma, d) \end{aligned}$$

is positive.

Proof By Proposition 3, $M^{\text{CH}(\gamma, d)}(K) = M^{[d]_0}([\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma])$. We choose $C^* > 0$ as in Proposition 5 for $[a, b] = [\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma]$ and we fix $\varepsilon := \varrho(\mathcal{I})/C^*$. Since $\text{dist}(M^{\text{CH}(\gamma, d)}, \mathbf{v}) > \varepsilon$, the vector \mathbf{v} does not belong to $N_\varepsilon(M^{\text{CH}(\gamma, d)})$. Hence, in view of Proposition 5,

$$\mathbf{v} \notin \text{conv} \left(\bigcup_{I \in \mathcal{I}} \Delta^I \right) = \text{conv} (\{ \Phi^I \mathbf{u}^i : I \in \mathcal{I}, i \in [d]_0 \}).$$

By separation theorems, there exists a vector $\mathbf{c} \in \mathbb{R}^{\text{CH}(\gamma, d)}$ with $\|\mathbf{c}\|_\infty \leq 1$ and $\delta \in \mathbb{R}$ such that $\langle \mathbf{c}, \mathbf{v} \rangle < \delta$ and $\langle \mathbf{c}, \mathbf{u} \rangle \geq \delta$ for all $\mathbf{u} \in \text{conv}(\bigcup_{I \in \mathcal{I}} \Delta^I)$. Hence \mathbf{c} and δ are feasible for (15) and their corresponding objective value is positive. \square

4.5 Axis Chain Pattern

We call chains that lie on a coordinate axis *axis chains*. That is, axis chains are chains γ with $|\text{supp}(\gamma)| = 1$. They are helpful to strengthen multilinear relaxation by introducing just n new patterns. See for example Figure 7 Configuration 2.

4.6 Shifting Patterns

To generate new patterns, we can utilise the following proposition.

Proposition 6 *Let $P \subseteq \mathbb{N}^n$ be a pattern with $\text{supp}(P) \neq [n]$ and $\eta \in \mathbb{N}^n$ a vector with $\text{supp}(\eta) \subseteq [n] \setminus \text{supp}(P)$. Then*

$$M^{\eta+P}(K) = \text{conv}(\mathbf{x}_{\min}^{\eta} M^P(K) \cup \mathbf{x}_{\max}^{\eta} M^P(K)).$$

Proof The assertion follows from

$$\begin{aligned} M^{\eta+P}(K) &= \text{conv}(\{(\mathbf{x}^{\eta+\beta})_{\beta \in P} : \mathbf{x} \in K\}) \\ &= \text{conv}(\{\mathbf{x}^{\eta}(\mathbf{x}^{\beta})_{\beta \in P} : \mathbf{x} \in K\}) \end{aligned}$$

and the observation that \mathbf{x}^{η} and \mathbf{x}^{β} have no common factor since $\text{supp}(\eta) \cap \text{supp}(\beta) = \emptyset$. Hence

$$\begin{aligned} \text{conv}(\{\mathbf{x}^{\eta}(\mathbf{x}^{\beta})_{\beta \in P} : \mathbf{x} \in K\}) &= \text{conv}\left(\bigcup_{\mathbf{y} \in K} \mathbf{y}^{\eta} \{(\mathbf{x}^{\beta})_{\beta \in P} : \mathbf{x} \in K\}\right) \\ &= \text{conv}\left(\bigcup_{\mathbf{y} \in K} \mathbf{y}^{\eta} M^P(K)\right) \\ &= \text{conv}(\mathbf{x}_{\min}^{\eta} M^P(K) \cup \mathbf{x}_{\max}^{\eta} M^P(K)). \quad \square \end{aligned}$$

Corollary 6 *In the setting of Proposition 6, the separation problem (SP) for the moment body $M^{\eta+P}(K)$ and a point $\mathbf{v} \in \mathbb{R}^{\eta+P}$ can be formulated as follows:*

$$\begin{aligned} &\text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\ &\text{for} && \mathbf{c} \in \mathbb{R}^{\eta+P} \text{ and } \delta \in \mathbb{R} \\ &\text{subject to} && \mathbf{x}_{\min}^{\eta} \sum_{\beta \in P} \mathbf{c}_{\eta+\beta} \mathbf{x}^{\beta} \geq \delta && \text{for all } \mathbf{x} \in K, \\ &&& \mathbf{x}_{\max}^{\eta} \sum_{\beta \in P} \mathbf{c}_{\eta+\beta} \mathbf{x}^{\beta} \geq \delta && \text{for all } \mathbf{x} \in K, \\ &&& \mathbf{c}_{\eta+\beta} \in [-1, 1] && \text{for all } \beta \in P. \end{aligned}$$

4.7 Shifted Chain Pattern

We apply the shifting procedure to chain patterns and generate a new pattern type. Let $d \in \mathbb{N}$ and $\gamma, \eta \in B$ with $\text{supp}(\gamma) \cap \text{supp}(\eta) = \emptyset$. We call $\eta + \text{CH}(\gamma, d)$ *shifted chain pattern*. Using Proposition 6 we can represent the moment body $M^{\eta + \text{CH}(\gamma, d)}(K)$ as the convex hull of $\mathbf{x}_{\min}^{\eta} M^{\text{CH}(\gamma, d)}(K)$ and $\mathbf{x}_{\max}^{\eta} M^{\text{CH}(\gamma, d)}(K)$. Using the notation $p_{\mathbf{c}}(t) := \sum_{i=0}^d \mathbf{c}_{\eta+i\gamma} t^i$ we can formulate analogous results for chains for shifted chains.

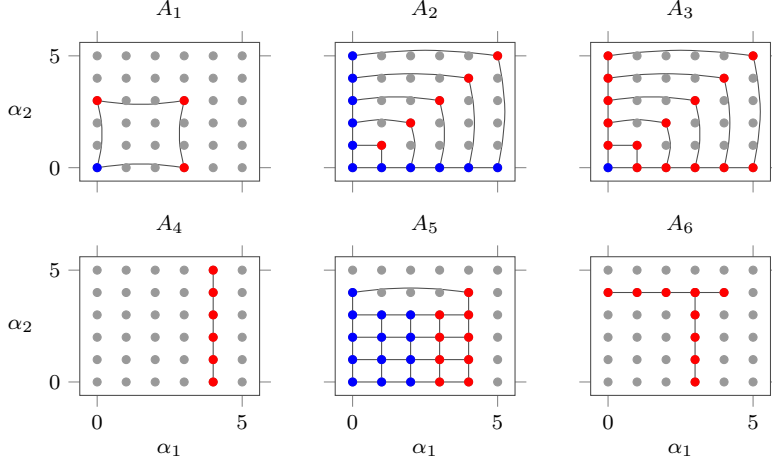


Fig. 5 Shifted chain patterns applied to A_1, \dots, A_6 ; formatting as in Figure 1.

Corollary 7 Let $\gamma, \eta \in \mathbb{N}^n \setminus \{\mathbf{0}\}$ have disjoint support. Let $g(t) = (\mathbf{x}_{\max}^{\gamma} - t)(t - \mathbf{x}_{\min}^{\gamma})$ and d be an even integer. Then problem (7) can be formulated as

$$\begin{aligned}
 & \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
 & \text{for} && \mathbf{c} \in \mathbb{R}^{\eta + \text{CH}(\gamma, d)}, \delta \in \mathbb{R}, \\
 & && \sigma_0, \tilde{\sigma}_0 \in \Sigma_{1, d} \text{ and } \sigma_1, \tilde{\sigma}_1 \in \Sigma_{1, d-2} \\
 & \text{subject to} && \mathbf{x}_{\min}^{\eta} p_{\mathbf{c}} - \delta = \sigma_0 + \sigma_1 g \\
 & && \mathbf{x}_{\max}^{\eta} p_{\mathbf{c}} - \delta = \tilde{\sigma}_0 + \tilde{\sigma}_1 g \\
 & && \mathbf{c}_{\beta + \eta} \in [-1, 1] \quad \text{for all } \beta \in \text{CH}(\gamma, d).
 \end{aligned}$$

Corollary 8 Let $\gamma, \eta \in \mathbb{N}^n \setminus \{\mathbf{0}\}$ have disjoint supports and let d be a positive integer. Then there exists a constant $C > 0$ that depends only on $\mathbf{x}_{\min}^{\gamma}, \mathbf{x}_{\max}^{\gamma}$ and d such that the following holds: If a vector $\mathbf{v} \in \mathbb{R}^{\eta + \text{CH}(\gamma, d)}$ does not belong to $M^{\eta + \text{CH}(\gamma, d)}(K)$ and \mathcal{I} is a covering of $[\mathbf{x}_{\min}^{\gamma}, \mathbf{x}_{\max}^{\gamma}]$ with

$$\varrho(\mathcal{I}) < C \text{dist}(M^{\eta + \text{CH}(\gamma, d)}(K), \mathbf{v}),$$

then the optimal value of the linear program

$$\begin{aligned}
& \text{maximize} && \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
& \text{for} && \mathbf{c} \in \mathbb{R}^{\eta + \text{CH}(\gamma, d)} \text{ and } \delta \in \mathbb{R} \\
& \text{subject to} && \sum_{i=0}^d \mathbf{c}_{\eta+i\gamma} \mathbf{w}_i \geq \delta && \text{for all } \mathbf{w} = \mathbf{x}_{\min}^\eta \Phi^I \mathbf{u}^i \text{ with} \\
& && && I \in \mathcal{I} \text{ and } i \in [d]_0, \\
& && \sum_{i=0}^d \mathbf{c}_{\eta+i\gamma} \mathbf{w}_i \geq \delta && \text{for all } \mathbf{w} = \mathbf{x}_{\max}^\eta \Phi^I \mathbf{u}^i \text{ with} \\
& && && I \in \mathcal{I} \text{ and } i \in [d]_0, \\
& && \mathbf{c}_\beta \in [-1, 1] && \text{for all } \beta \in \eta + \text{CH}(\gamma, d)
\end{aligned} \tag{16}$$

is positive.

Proof Let C^* be the constant from Proposition 5 for $[a, b] = [\mathbf{x}_{\min}^\gamma, \mathbf{x}_{\max}^\gamma]$ and $\kappa := \max\{|\mathbf{x}_{\min}^\eta|, |\mathbf{x}_{\max}^\eta|\}$. Using Proposition 6 we write

$$M^{\eta + \text{CH}(\gamma, d)}(K) = \text{conv}(\mathbf{x}_{\min}^\eta M^{\text{CH}(\gamma, d)}(K) \cup \mathbf{x}_{\max}^\eta M^{\text{CH}(\gamma, d)}(K)).$$

We claim that

$$\mathbf{x}_{\min}^\eta N_\varepsilon(M^{\text{CH}(\gamma, d)}(K)) \cup \mathbf{x}_{\max}^\eta N_\varepsilon(M^{\text{CH}(\gamma, d)}(K))$$

is a subset of

$$N_{\kappa\varepsilon}(\text{conv}(\mathbf{x}_{\min}^\eta M^{\text{CH}(\gamma, d)}(K) \cup \mathbf{x}_{\max}^\eta M^{\text{CH}(\gamma, d)}(K))).$$

Once the claim is established, the assertion follows with $C = \frac{C^*}{\kappa}$ using the same arguments as in the proof of Corollary 5.

For the proof of the claim let $\mathbf{v} \in \mathbf{x}_{\min}^\eta N_\varepsilon(M^{\text{CH}(\gamma, d)}(K))$. Then there exists $\mathbf{u} \in M^{\text{CH}(\gamma, d)}(K)$ with $\|\frac{\mathbf{v}}{|\mathbf{x}_{\min}^\eta|} - \mathbf{u}\|_1 \leq \varepsilon$. Hence $\|\mathbf{v} - \mathbf{x}_{\min}^\eta \mathbf{u}\|_1 \leq |\mathbf{x}_{\min}^\eta| \varepsilon$ and therefore

$$\mathbf{v} \in N_{|\mathbf{x}_{\min}^\eta| \varepsilon}(\mathbf{x}_{\min}^\eta M^{\text{CH}(\gamma, d)}(K)).$$

Analogously, if $\mathbf{v} \in \mathbf{x}_{\max}^\eta N_\varepsilon(M^{\text{CH}(\gamma, d)}(K))$, then

$$\mathbf{v} \in N_{|\mathbf{x}_{\max}^\eta| \varepsilon}(\mathbf{x}_{\min}^\eta M^{\text{CH}(\gamma, d)}(K)),$$

which concludes the proof of the claim. \square

5 Expressing Known Convexification Techniques from our Viewpoint

5.1 Expression Trees

Covexification using expression trees is common in general nonlinear optimization [28]. This approach is based on the observation that each algebraic expression is made up of a certain set of elementary operations. In the context of polynomial optimization, as elementary operations one could choose, for example, taking a power of a term, taking a linear combination and taking a product of terms.

A decomposition of an algebraic expression into these operations can be visualized using an algebraic expression tree, like in Figure 6. This is a rooted tree with nodes labeled by terms occurring in the expression. Each term is built up from its child terms using elementary operations and the underlying convexification is obtained by introducing a variable for each node and providing convex constraints that link every node and its child nodes. For polynomials, given as a linear combination of monomials, all the nodes apart from the root node correspond to monomial variables. A non-root node and its child nodes therefore build a pattern.

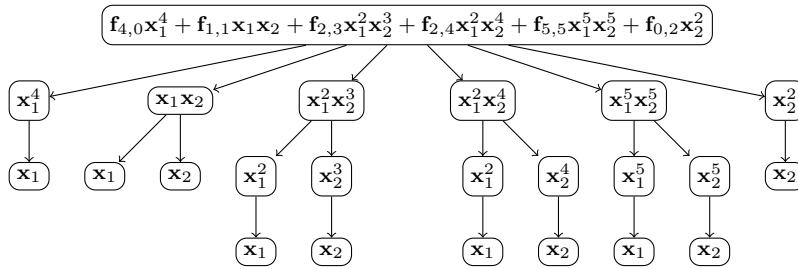


Fig. 6 A possible algebraic expression tree for the polynomial f^{ex1} and the set A from Example 1.

For example, the term $x_1^2x_2^3$ in Figure 6 is decomposed into the product of the powers x_1^2 and x_2^3 of the variables x_1 and x_2 . For these three terms, one introduces the monomial variables $\mathbf{v}_{(2,3)}$, $\mathbf{v}_{(2,0)}$ and $\mathbf{v}_{(0,3)}$, respectively. The relationship of these variables is captured by the pattern

$$P = \{(2, 3), (2, 0), (0, 3)\}$$

and the corresponding moment body $M^P(K)$ is described by the well-known McCormick inequalities. The variable $\mathbf{v}_{(0,3)}$ is further connected to $\mathbf{v}_{(0,1)}$ by exponentiation. The corresponding pattern is $\{(0, 1), (0, 3)\}$.

Since expression trees normally correspond to patterns of small size, they lead to weak but efficiently computable relaxations. Such relaxations are then employed within the branch-and-bound framework in order to compensate the

poor lower bounds. The computational costs of such strategies strongly depend on the quality of the generated lower bounds. If the underlying bounds are too weak, the branch-and-bound based approach is not computationally feasible.

5.2 Bound-Factor Products

Another common convexification approach from general nonlinear optimization is based on so-called *bound-factor products* [13]. Since the polynomials $\mathbf{x}_i - \mathbf{a}_i$ and $\mathbf{x}_i - \mathbf{x}_i$ are nonnegative on K , the products of these polynomials (with repetitions allowed) are also nonnegative on K . So, one can consider the products

$$F^{\alpha,\beta}(\mathbf{x}) := \prod_{i=1}^n (\mathbf{x}_i - \mathbf{a}_i)^{\alpha_i - \beta_i} (\mathbf{b}_i - \mathbf{x}_i)^{\beta_i} \quad (17)$$

of $|\alpha|$ polynomials with α_i linear factors depending on the variable \mathbf{x}_i , where $\alpha, \beta \in \mathbb{N}^n$ and $\alpha \geq \beta$. For a generic choice of \mathbf{a} and \mathbf{b} , the polynomial $F^{\alpha,\beta}(\mathbf{x})$ includes all monomials with exponents in the pattern $\text{BF}(\alpha) := \{0, \dots, \alpha_1\} \times \dots \times \{0, \dots, \alpha_n\}$. By substituting $\mathbf{v}_\gamma = \mathbf{x}^\gamma$ for all $\gamma \in \text{BF}(\alpha)$ we obtain a linearization $LF^{\alpha,\beta}(\mathbf{v})$ of $F^{\alpha,\beta}(\mathbf{x})$. The system of linear inequalities

$$LF^{\alpha,\beta}(\mathbf{v}) \geq 0 \text{ for all } \beta \in \text{BF}(\alpha) \quad (18)$$

is valid for $\mathbf{v} \in M^{\text{BF}(\alpha)}(K)$. Thus, within this approach one groups monomial variables into patterns of a rather big size and connects them with only linear constraints. For example, to generate a non-trivial relaxation of (POP) using bound-factor products for the set A from Example 1 one is forced to use at least one pattern $\text{BF}(\alpha)$ with $\alpha_1 \geq 5$ and $\alpha_2 \geq 5$, which means that at least 36 monomial variables have to be introduced. Another issue is that the system of linear inequalities (18) is not a tight description of $M^{\text{BF}(\alpha)}(K)$. These kinds of relaxations have also been used within branch-and-bound strategies.

5.3 Moment Relaxation

The most popular convexification techniques in the polynomial optimization community are *moment relaxations* and their dual counterparts *sum-of-squares relaxations* [5, 21, 22]. This approach introduces a large number of monomial variables and links them all within one large group using semidefinite constraints. The approach is hierarchical in the sense that one first needs to choose a bound on the degree of the monomials, for which monomial variables are introduced. These hierarchies have good approximation properties at the expense of large SDPs. Even the lowest possible hierarchy level of the moment relaxation for medium-sized problems results in an absurdly huge SDP. However, strategies exist to make this approach more tractable, e.g., [3] and [20, Ch. 8].

To derive a so-called moment relaxation of (POP), the following representation of the moment body $M^A(K)$ in terms of probability measures is used:

$$M^A(K) = \left\{ \int m^A(\mathbf{x})\mu(d\mathbf{x}) : \mu \text{ is a probability measure with } \text{supp}(\mu) \subseteq K \right\}.$$

So a vector $\mathbf{v} \in \mathbb{R}^A$ belongs to $M^A(K)$ if and only if there exists a probability measure μ with $\text{supp}(\mu) \subseteq K$ such that $\mathbf{v}_\alpha = \int \mathbf{x}^\alpha \mu(d\mathbf{x})$ for all $\alpha \in A$. Hence, (C-POP) can be formulated as

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\mathbb{N}^n} \\ & \text{subject to} && \mathbf{v} \text{ is a moment sequence of a probability measure on } K. \end{aligned}$$

In order to obtain a tractable characterization of the feasible set, we use the following definition and theorem.

Definition 1 (Moment Matrix and Localizing Matrix [20, Ch.2.7.1])

The localizing matrix $M(g, \mathbf{v})$ for a polynomial g with coefficients $(\mathbf{g}_\alpha)_\alpha$ and the moment matrix $\mathbf{M}_k(\mathbf{v})$ are defined as

$$\begin{aligned} \mathbf{M}_k(g, \mathbf{v}) &:= \left(\sum_{\gamma \in \mathbb{N}^n} \mathbf{g}_\gamma \mathbf{v}_{\gamma+\alpha+\beta} \right)_{\alpha, \beta \in \mathbb{N}_k^n} \\ \mathbf{M}_k(\mathbf{v}) &:= \mathbf{M}_k(1, \mathbf{v}). \end{aligned}$$

Theorem 3 ([20, Th. 2.44]) *Let $g_i, i \in [m]$, be n -variate polynomials such that there exist sum-of-squares polynomials $\sigma_i, i \in [m]_0$, for which*

$$\{\mathbf{x} \in \mathbb{R}^n : \sigma_0(\mathbf{x}) + \sum_{i \in [m]} \sigma_i(\mathbf{x})g_i(\mathbf{x}) \geq 0\}$$

is compact. Furthermore, let

$$K = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \geq 0, i \in [m]\}.$$

A sequence $(\mathbf{v}_\alpha)_\alpha$ has a finite Borel representing measure with support in K if and only if

$$\begin{aligned} & \mathbf{M}_k(\mathbf{v}) \text{ is psd} \\ & \mathbf{M}_k(g_j, \mathbf{v}) \text{ is psd for all } j \in [m] \end{aligned}$$

for all k .

We describe the box K by the polynomials $g_i(\mathbf{x}) := (\mathbf{x}_i - \mathbf{a}_i)(\mathbf{b}_i - \mathbf{x}_i)$ for $i \in [n]$, i.e. $K = \{\mathbf{x} \in \mathbb{R}^n : g_i(\mathbf{x}) \geq 0, i \in [n]\}$. Clearly, the assumptions of Theorem 3 hold and we can formulate (C-POP) as

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\mathbb{N}^n} \\ & \text{subject to} && \mathbf{M}_k(\mathbf{v}) \quad \text{is psd for all } k, \\ & && \mathbf{M}_k(g_i, \mathbf{v}) \text{ is psd for all } i \in [n] \text{ and all } k, \\ & && \mathbf{v}_0 = 1. \end{aligned} \tag{19}$$

By truncating the infinite dimensional matrices we obtain a finite-dimensional problem. For every $d \geq \lceil \frac{\deg(A)}{2} \rceil$ the optimal value ρ_d of the semidefinite problem

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\mathbb{N}_{2d}^n} \\ & \text{subject to} && \mathbf{M}_d(\mathbf{v}) \quad \text{is psd,} \\ & && \mathbf{M}_{d-2}(g_i, \mathbf{v}) \text{ is psd for all } i \in [n], \\ & && \mathbf{v}_0 = 1 \end{aligned} \tag{20}$$

is a lower bound on the optimal value of (POP). This problem has one SDP constraint of size $\binom{n+d}{d}$ that involves the variables $\mathbf{v}_\alpha, \alpha \in \mathbb{N}_{2d}^n$, and n SDP constraints of size $\binom{n+d-1}{d-1}$ that involve $\mathbf{v}_\alpha, \alpha \in \mathbb{N}_{2d-2}^n$. Note that for general problems it is not possible to reduce the size of the mentioned SDP constraints [7]. For a small set A of degree 10 and $n = 2$ like in Example 1 this already adds up to 66 moment variables. The subfigure for the moment relaxation in Figure 1 shows the biggest SDP constraint with a 66×66 matrix.

5.4 Alternative Convexification Techniques

Apart from the mentioned techniques, alternative convexification approaches useful in the context of polynomial optimization and based on geometric and signomial programming have been investigated in [17, 18, 12, 11]. Closely related to geometric and signomial programming are so-called SONC Positivstellensatz [19, 15, 27, 16]. By dualizing the SONC relaxations, one arrives at convexifications in terms of monomial variables [16].

5.5 Pattern Approach

The pattern approach enables us to adjust the size and tractability of a relaxation. In Figure 7 we give 3 different pattern configurations. Configuration 1 involves 23 moment variables, which are connected by 2 multilinear patterns, 3 chains and 2 shifted chains. Configuration 2 involves 15 moment variables.

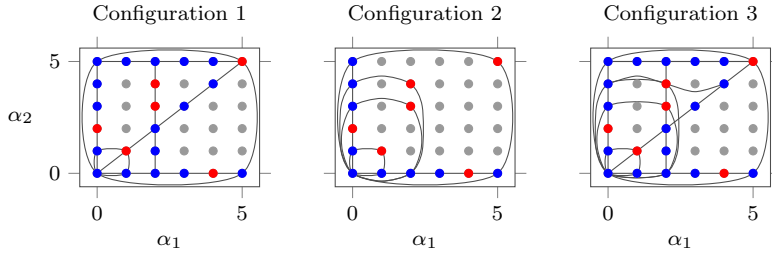


Fig. 7 Three different manually generated pattern configurations applied to the set A from Example 1.

These variables are connected by the same 4 multilinear patterns that are used in the tree reformulation relaxation depicted in Figure 1. But in contrast to the tree reformulation we use the 2 axis chains

$$\begin{aligned} \text{CH}(\mathbf{e}^1, 5) &= \{(0, 0), (1, 0), (2, 0), (3, 0), (4, 0), (5, 0)\}, \\ \text{CH}(\mathbf{e}^2, 5) &= \{(0, 0), (0, 1), (0, 2), (0, 3), (0, 4), (0, 5)\} \end{aligned}$$

instead of the seven smaller patterns

$$\begin{aligned} &\{(1, 0), (2, 0)\}, \{(1, 0), (4, 0)\}, \{(1, 0), (5, 0)\}, \\ &\{(0, 1), (0, 2)\}, \{(0, 1), (0, 3)\}, \{(0, 1), (0, 4)\}, \{(0, 1), (0, 5)\}. \end{aligned}$$

Therefore, the chain $\text{CH}(\mathbf{e}^2, 5)$ connects $\mathbf{v}_{(2,0)}$, $\mathbf{v}_{(3,0)}$, $\mathbf{v}_{(4,0)}$ and $\mathbf{v}_{(5,0)}$ directly, whereas in the tree reformulation those variables are only indirectly connected. Configuration 3 involves the same number of moment variables as Configuration 1, but provides a better connection of these variables through additional patterns.

6 Algorithm

We present a cutting-plane algorithm that iteratively solves (P-RLX) and suggest a method for choosing a system of patterns for relaxing (POP) to (P-RLX).

6.1 Cutting-Plane Algorithm

Let $\varepsilon > 0$. Considering an optimization problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in X\}. \quad (21)$$

with the optimal value f^* we say that $\mathbf{x} \in \mathbb{R}^n$ is ε -feasible for (21) if $\mathbf{x} \in N_\varepsilon(X)$ and ε -optimal for (21) if

$$f^* - \varepsilon \leq f(\mathbf{x}) \leq f^* + \varepsilon.$$

Input: $\mathbf{f} \in \mathbb{R}^{\bar{A}}$, finite family of patterns $\mathcal{P} := \{P_1, \dots, P_m\}$ with $\bar{A} = P_1 \cup \dots \cup P_m$ and $\varepsilon > 0$.

Output: $O(\varepsilon)$ -optimal and $\mathcal{O}(\varepsilon)$ -feasible $\mathbf{v}^*(\varepsilon)$ of (P-RLX).

- (0) Solve LP corresponding to the pattern relaxation of $M^{\bar{A}}(K)$ with just singleton patterns, i.e.

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\bar{A}} \\ & \text{subject to} && \mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha] \quad \text{for all } \alpha \in \bar{A}. \end{aligned}$$

Let $\mathbf{v}^0 \in \mathbb{R}^{\bar{A}}$ be the optimal point and Ineq^0 be the set of inequalities $\mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha]$ for all $\alpha \in \bar{A}$. Set $i = 1$.

- (1) Initialize $\text{Ineq}^i = \emptyset$. For each pattern P_j , $j \in [m]$, determine the distance $\text{dist}(M^{P_j}, \mathbf{v}_{P_j})$ by solving the respective separation problem (SP). If $\text{dist}(M^{P_j}, \mathbf{v}_{P_j}) > \varepsilon$, add the inequality $\langle \mathbf{c}^j, \mathbf{v} \rangle \geq \delta$, $\mathbf{v} \in \mathbb{R}^{\bar{A}}$ to the set of inequalities Ineq^i .
- (2) If the set Ineq^i is empty, return $\mathbf{v}^*(\varepsilon) := \mathbf{v}^i$.
- (3) Solve the auxiliary problem

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\bar{A}} \\ & \text{subject to} && \mathbf{v} \text{ satisfies } \text{Ineq}^k \text{ for all } k = 0, \dots, i \end{aligned} \tag{22}$$

and save the minimizer as \mathbf{v}^{i+1} .

- (4) Set $i \leftarrow i + 1$ and go to step (1).

Theorem 4 For every given $\varepsilon > 0$, the Cutting-Plane Algorithm 6.1 terminates after a finite number of iterations. The output satisfies

$$\mathbf{v}_{P_j}^*(\varepsilon) \in N_\varepsilon(M^{P_j}(K)) \quad \text{for all } j \in [m]. \tag{23}$$

Proof Assume that the Cutting-Plane Algorithm 6.1 does not terminate after a finite number of iterations. Then it produces an infinite sequence $(\mathbf{v}^i)_{i \in \mathbb{N}}$ such that for all i there exists a $j(i) \in [m]$ with

$$\mathbf{v}_{P_{j(i)}} \notin N_\varepsilon(M^{P_{j(i)}}(K)).$$

Hence, there exists a pattern P and infinite sequence $(i_k)_{k \in \mathbb{N}}$ satisfying $P = P_{j(i_k)}$ for all k . Let F^i be the feasible set of (22) in the i -th iteration. Observe that by construction $F^{i+1} \subseteq F^i$ holds and therefore $\mathbf{v}^{i_k} \in F^0$ for all $k \in \mathbb{N}$. Since F^0 is compact there exists a converging subsequence $(\tilde{\mathbf{v}}^i)_{i \in \mathbb{N}}$ of $(\mathbf{v}^{i_k})_{k \in \mathbb{N}}$. Let $\tilde{\mathbf{v}} = \lim_{i \rightarrow \infty} \tilde{\mathbf{v}}^i$. By the choice of the sequence we have

$$\text{dist}(M^P(K), \tilde{\mathbf{v}}_P^i) > \varepsilon.$$

for all $i \in \mathbb{N}$. Hence, for i large enough, $\|\tilde{\mathbf{v}}_P^i - \tilde{\mathbf{v}}_P\|_1 < \text{dist}(M^P(K), \tilde{\mathbf{v}}_P^i)$ holds. Application of Proposition 1 to the minimizers \mathbf{c}^i, δ_i of the problem (SP) in the case $\mathbf{v} = \tilde{\mathbf{v}}_P^i$ yields

$$\begin{aligned} \langle \mathbf{c}^i, \tilde{\mathbf{v}}_P \rangle &= \langle \mathbf{c}^i, \tilde{\mathbf{v}}_P - \tilde{\mathbf{v}}_P^i \rangle + \langle \mathbf{c}^i, \tilde{\mathbf{v}}_P^i \rangle \\ &= \langle \mathbf{c}^i, \tilde{\mathbf{v}}_P - \tilde{\mathbf{v}}_P^i \rangle + \delta_i - \text{dist}(M^P(K), \tilde{\mathbf{v}}_P^i) \\ &< \text{dist}(M^P(K), \tilde{\mathbf{v}}_P^i) + \delta_i - \text{dist}(M^P(K), \tilde{\mathbf{v}}_P^i) = \delta_i. \end{aligned}$$

This is a contradiction since $\tilde{\mathbf{v}} \in F^i$ for all i . \square

The following theorem shows that $f^*(\varepsilon) := \langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle$ converges to the optimal value f^* of (P-RLX), as $\varepsilon \rightarrow 0$, and that the convergence rate depends linearly on ε .

Theorem 5 *There exists a constant $C(K, \bar{A})$ depending on K and \bar{A} such that for every $\varepsilon > 0$ the distance between the feasible set of (P-RLX) and the output $\mathbf{v}^*(\varepsilon)$ of the Cutting-Plane Algorithm 6.1 is at most $C(K, \bar{A})\varepsilon$. Furthermore, the optimal value of (P-RLX) f^* satisfies*

$$\langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle \leq f^* \leq \|\mathbf{f}\|_\infty C(K, \bar{A})\varepsilon + \langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle.$$

For the proof of the proposition we use the following lemma, which is an adaption of [25, Lem. 1.8.9].

Lemma 2 *Let X, X^1, \dots, X^m be nonempty compact convex subsets of \mathbb{R}^A such that the intersection $Y := X^1 \cap \dots \cap X^m$ contains an l_1 -ball of radius $\rho > 0$ as a subset and the inclusion $X^i \subseteq X$ holds for every $i \in [m]$. Let $\varepsilon > 0$ and let \mathbf{x} be a point of X satisfying $\text{dist}(X^i, \mathbf{x}) \leq \varepsilon$ for every $i \in [m]$. Then*

$$\text{dist}(Y, \mathbf{x}) \leq \frac{\varepsilon}{\rho} \text{diam}(X).$$

Proof Since the assertion is invariant under translations, we assume that the l_1 -ball of radius ρ contained in Y is centered at the origin, that is, $B := \{x \in \mathbb{R}^A : \|x\|_1 \leq \rho\} \subseteq Y$. For each $i \in [m]$, choose a point $\mathbf{x}^i \in X^i$ with $\|\mathbf{x} - \mathbf{x}^i\|_1 \leq \varepsilon$. We claim that the point $\mathbf{y} := \frac{\rho}{\rho + \varepsilon} \mathbf{x}$ belongs to Y . For every $i \in [m]$, we fix $\mathbf{p}^i \in \mathbb{R}^A$ defined by the equality

$$\mathbf{y} = \frac{\rho}{\rho + \varepsilon} \mathbf{x} = \frac{\varepsilon}{\rho + \varepsilon} \mathbf{p}^i + \frac{\rho}{\rho + \varepsilon} \mathbf{x}^i.$$

By construction, \mathbf{y} is a convex combination of \mathbf{p}^i and \mathbf{x}^i . Thus, for verifying the claim, it suffices to show $\mathbf{p}^i \in B$ for every $i \in [m]$. Indeed, if $\mathbf{p}^i \in B$, then since B is a subset of X^i and \mathbf{p}^i belongs to X^i , we obtain $\mathbf{y} \in X^i$ for every $i \in [m]$, which verifies the claim. The point \mathbf{p}^i can be defined explicitly as

$$\mathbf{p}^i = \frac{\rho}{\varepsilon} (\mathbf{x} - \mathbf{x}^i).$$

Since $\|\mathbf{x} - \mathbf{x}^i\| \leq \varepsilon$, we immediately get $\mathbf{p}^i \in B$. The proof is concluded by estimating the distance between \mathbf{x} and \mathbf{y} :

$$\text{dist}(Y, \mathbf{x}) \leq \|\mathbf{y} - \mathbf{x}\|_1 = \frac{\varepsilon}{\rho + \varepsilon} \|\mathbf{x}\|_1.$$

Here, $\frac{\varepsilon}{\rho + \varepsilon} \leq \frac{\varepsilon}{\rho}$, while $\|\mathbf{x}\|_1$ is the l_1 -distance between $\mathbf{0}$ and \mathbf{x} both belonging to X , which implies $\|\mathbf{x}\|_1 \leq \text{diam}(X)$. Thus, we arrive at a desired estimate of $\text{dist}(Y, \mathbf{x})$. \square

Proof (Theorem 5) If some pattern P_i contains $\mathbf{0}$, then $\mathbf{0} \in \bar{A}$ and the constraint $\mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha]$ with $\alpha = \mathbf{0}$ occurring in Ineq^0 can be formulated as the equality $\mathbf{v}_\mathbf{0} = 1$. This shows that all solutions generated during the iteration satisfy the constraint $\mathbf{v}_\mathbf{0} = 1$. Using this observation and Theorem 4, which yields $\mathbf{v}_{P_i}^*(\varepsilon) \in N_\varepsilon(M^{P_i}(K))$, we obtain $\mathbf{v}_{P_i \setminus \{\mathbf{0}\}}^*(\varepsilon) \in N_\varepsilon(M^{P_i \setminus \{\mathbf{0}\}}(K))$. We can therefore remove $\mathbf{0}$ from all patterns and assume that the patterns $\mathbf{0} \notin P_i$ holds for every $i \in [m]$.

Let

$$F := \{\mathbf{v} \in \mathbb{R}^{\bar{A}} : \mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha] \text{ for all } \alpha \in \bar{A}\}.$$

The feasible set of (P-RLX) is the intersection of

$$F^i := \{\mathbf{v} \in \mathbb{R}^{\bar{A}} : \mathbf{v}_{P_i} \in M^{P_i}(K)\} \cap F$$

for all $i \in [m]$. We show that $M^{\bar{A}}(K)$ is full dimensional by assuming the contrary. Then $M^{\bar{A}}(K)$ is contained in a linear subspace $\{\mathbf{v} \in \mathbb{R}^{\bar{A}} : \langle \mathbf{c}, \mathbf{v} \rangle = 0\}$ given by $\mathbf{c} \in \mathbb{R}^{\bar{A}}$ with $\mathbf{c} \neq \mathbf{0}$. Hence $\langle \mathbf{c}, m^{\bar{A}}(\mathbf{x}) \rangle$ is a polynomial in \mathbf{x} vanishing on a n -dimensional set K . This implies $\mathbf{c} = \mathbf{0}$, which is a contradiction. Hence $M^{\bar{A}}(K)$ is full dimensional and therefore contains a ball with radius $R(K, \bar{A})$ depending only on K and \bar{A} . The feasible set $\bigcap_{i \in [m]} F^i \cap F$ of (P-RLX) contains $M^{\bar{A}}(K)$ and therefore the mentioned ball. The diameter $D(K, \bar{A})$ of F depends only on K and \bar{A} . From $\mathbf{v}^*(\varepsilon) \in \bigcap_{i \in [m]} N_\varepsilon(F^i) \cap F$ and Lemma 2 follows

$$\text{dist}\left(\bigcap_{i \in [m]} (F^i \cap F), \mathbf{v}^*(\varepsilon)\right) \leq \underbrace{\frac{D(K, \bar{A})}{R(K, \bar{A})}}_{:=C(K, \bar{A})} \varepsilon.$$

At last, let $\mathbf{v} \in \bigcap_{i \in [m]} (F^i \cap F)$ satisfying $\|\mathbf{v} - \mathbf{v}^*(\varepsilon)\|_1 \leq C(K, \bar{A})\varepsilon$. The upper bound on f^* in terms of $\mathbf{v}^*(\varepsilon)$ follows from

$$\begin{aligned} f^* &\leq \langle \mathbf{f}, \mathbf{v} \rangle \\ &= \langle \mathbf{f}, \mathbf{v} - \mathbf{v}^*(\varepsilon) \rangle + \langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle \\ &\leq \|\mathbf{f}\|_\infty \|\mathbf{v} - \mathbf{v}^*(\varepsilon)\|_1 + \langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle \\ &= \|\mathbf{f}\|_\infty C(K, \bar{A})\varepsilon + \langle \mathbf{f}, \mathbf{v}^*(\varepsilon) \rangle. \end{aligned} \quad \square$$

If the separation problem for some patterns P is too hard, it might be useful to replace the corresponding moment bodies $M^P(K)$ by convex and compact approximations $M_{\varepsilon'}^P$, $\varepsilon' > 0$, that satisfy

$$M^P(K) \subseteq M_{\varepsilon'}^P \subseteq N_{\varepsilon'}(M^P(K)).$$

Since the linear constraints $\mathbf{x}_{\min}^\alpha \leq \mathbf{v}_\alpha \leq \mathbf{x}_{\max}^\alpha$ with $\alpha \in P$ are valid for $M^P(K)$, one can always add these constraints to the underlying approximate description of $M^P(K)$. We can therefore assume that $M_{\varepsilon'}^P$ is a subset of the box $\{\mathbf{v} \in \mathbb{R}^P : \mathbf{v}_\alpha \in [\mathbf{x}_{\min}^\alpha, \mathbf{x}_{\max}^\alpha], \text{ for all } \alpha \in P\}$. Replacing the separation problems in step (1) of the Cutting-Plane Algorithm 6.1 by the separation problems for $M_{\varepsilon'}^{P_i}$ yields an algorithm that solves

$$\begin{aligned} & \text{minimize} && \langle \mathbf{f}, \mathbf{v} \rangle \\ & \text{for} && \mathbf{v} \in \mathbb{R}^{\bar{A}} && (\varepsilon'\text{-P-RLX}) \\ & \text{subject to} && \mathbf{v}_{P_i} \in M_{\varepsilon'}^{P_i} \text{ for all } i \in [m]. \end{aligned}$$

This algorithm terminates after finitely many iterations. To see, it suffices to observe that Theorem 4 holds for $M_{\varepsilon'}^{P_j}$ in place of $M^{P_j}(K)$, since the proof of Theorem 4 only relies on the convexity and compactness of $M^{P_j}(K)$. Similarly, the proof of Theorem 5 can be used without any changes to show that the optimal value $f^{\varepsilon'}$ of $(\varepsilon'\text{-P-RLX})$ and the output $\mathbf{v}^{\varepsilon'}(\varepsilon)$ of the algorithm satisfy

$$\langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon) \rangle \leq f^{\varepsilon'} \leq \|\mathbf{f}\|_\infty C(K, \bar{A})\varepsilon + \langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon) \rangle. \quad (24)$$

Since every $(\varepsilon'\text{-P-RLX})$ feasible point is $C(K, \bar{A})\varepsilon$ -feasible for (P-RLX), we have

$$f^* \leq \|\mathbf{f}\|_\infty C(K, \bar{A})\varepsilon' + f^{\varepsilon'}. \quad (25)$$

Combining (24) and (25) using the triangle inequality yields

$$\langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon) \rangle \leq f^* \leq \|\mathbf{f}\|_\infty C(K, \bar{A})(\varepsilon + \varepsilon') + \langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon) \rangle.$$

This line of thought justifies replacing separation problems for chains and shifted chains in step (1) of Cutting-Plane Algorithm 6.1 by (15) and (16).

6.2 Pattern Generation Routine

For being able to use Cutting-Plane Algorithm 6.1, we need to develop a way of generating a family of patterns for a given nonempty set $A \subseteq \mathbb{N}^n$. The following is an algorithm for generation of a family of patterns \mathcal{P} for A and some given pattern types $\text{Type}_1, \dots, \text{Type}_k$. To use this algorithm, we need to supply each pattern type Type_i with a method, which we call Find Pattern Routine, that takes a set $\bar{A} \subseteq \mathbb{N}^n$ as an input and generates a family of patterns of Type_i connecting the elements of \bar{A} . In what follows we make suggestions for Find Pattern Routine methods for the pattern types that we described in Section 4.

Input: $A \subseteq \mathbb{N}^n$ and pattern types $\text{Type}_1, \dots, \text{Type}_k$.
Output: a family \mathcal{P} of patterns of types $\text{Type}_1, \dots, \text{Type}_k$ and a set $\bar{A} \subseteq \mathbb{N}^n$ with $A \subseteq \bar{A}$ and $\bar{A} = \bigcup_{P \in \mathcal{P}} P$.

- (1) Initialize
 $\bar{A} \leftarrow A$
 $\mathcal{P} \leftarrow \emptyset$.
- (2) For all $i \in [k]$
employ Find Pattern Routine for Type_i and obtain family of patterns \mathcal{P}_i of Type_i and updated \bar{A} .
- (3) Return the family of patterns $\mathcal{P} = \bigcup_{i \in [k]} \mathcal{P}_i$ and \bar{A} .

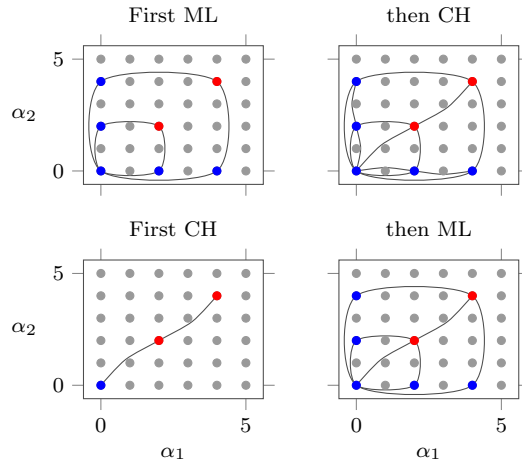


Fig. 8 For the given set A the Pattern Generation Routine 6.2 produces different results for two different orderings of the pattern types ML and CH.

The output of the Pattern Generation Routine 6.2 depends on the order of the pattern types $\text{Type}_1, \dots, \text{Type}_k$, see Figure 8. This is because the set \bar{A} grows due to its update in the Add Pattern Routine. For dense polynomials of degree d , which correspond to $A = \mathbb{N}_d^n$, the order of the pattern types does not matter.

6.3 Add Pattern Routine

This routine adds a pattern P to a given family \mathcal{P} if P is not already contained in a pattern of \mathcal{P} .

Input: $\bar{A} \subseteq \mathbb{N}^n$, a family of patterns \mathcal{P} and a pattern P .
Output: Updated \mathcal{P} and \bar{A} .

If $P \not\subseteq \tilde{P}$ for all $\tilde{P} \in \mathcal{P}$
 $\mathcal{P} \leftarrow \mathcal{P} \cup \{P\}$
 $\bar{A} \leftarrow \bar{A} \cup P.$

6.4 Find Multilinear Pattern Routine

Input: $\bar{A} \subseteq \mathbb{N}^n$.
Output: a family \mathcal{P} of multilinear patterns and a set $\bar{A} \subseteq \mathbb{N}^n$ with $A \subseteq \bar{A}$ and $\bar{A} = \bigcup_{P \in \mathcal{P}} P$.

- (1) Initialize
 $\mathcal{P} \leftarrow \emptyset.$
- (2) For $\alpha \in \bar{A}$
use Add Pattern Routine 6.3 for $\text{ML}(\alpha)$ to update \mathcal{P} and \bar{A} .
- (3) Return the family of patterns \mathcal{P} and \bar{A} .

6.5 Find Chain Pattern Routine

As usual, $\text{gcd}(\alpha)$ is the greatest common divisor of the components of $\alpha \in \mathbb{N}^n$. For $X \subseteq \mathbb{N}$, $\text{gcd}(X)$ denotes the greatest common divisor of the elements of X .

Input: $\bar{A} \subseteq \mathbb{N}^n$.
Output: a family \mathcal{P} of chain patterns and a set $\bar{A} \subseteq \mathbb{N}^n$ with $A \subseteq \bar{A}$ and $\bar{A} = \bigcup_{P \in \mathcal{P}} P$.

- (1) Initialize
 $\mathcal{P} \leftarrow \emptyset.$
- (2) For all $\beta \in \{\frac{\alpha}{\text{gcd}(\alpha)} : \alpha \in \bar{A} \setminus \{\mathbf{0}\}\}$,
 - (2.1) let s be the maximal integer value such that $s\beta \in \bar{A}$ and
 $g = \text{gcd}(\{t \in \mathbb{N} : t\beta \in \bar{A}\})$,
 - (2.2) then use the Add Pattern Routine 6.3 for
$$\text{CH}(g\beta, \frac{s}{g}) = \left\{ ig\beta : i = 0, \dots, \frac{s}{g} \right\}$$

to update \mathcal{P} and \bar{A} .
- (3) Return the family of patterns \mathcal{P} and \bar{A} .

In step (2.2), we could also add the chain $\text{CH}(\beta, s)$ instead of $\text{CH}(g\beta, \frac{s}{g})$, since $\text{CH}(g\beta, \frac{s}{g}) \subseteq \text{CH}(\beta, s)$. When $\text{CH}(g\beta, \frac{s}{g})$ is generated in step (2.2), it connects the same moment variables in \bar{A} as the chain $\text{CH}(\beta, s)$. In our algorithm, we prefer to use $\text{CH}(g\beta, \frac{s}{g})$, because in general this chain has a smaller cardinality and by this it introduces fewer new moment variables.

6.6 Find Shifted Chain Pattern Routine

This routine finds shifted chains with generators $\gamma = \mathbf{e}^i, i \in [n]$. We define the projection $\pi^i(\alpha) := (\alpha_1, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_n)$ of a vector $\alpha \in \mathbb{N}^n$.

Input: $\bar{A} \subseteq \mathbb{N}^n$.

Output: a family \mathcal{P} of shifted chain patterns with generators $\gamma = \mathbf{e}^i, i \in [n]$ and a set $\bar{A} \subseteq \mathbb{N}^n$ with $A \subseteq \bar{A}$ and $\bar{A} = \bigcup_{P \in \mathcal{P}} P$,

(1) Initialize

$$\mathcal{P} \leftarrow \emptyset.$$

(2) For all $i \in [n]$

for all $\beta \in \pi^i(\bar{A})$

(2.1) let $s = \max\{\alpha_i : \alpha \in \bar{A}, \pi^i(\alpha) = \beta\}$,

$g = \gcd(\{\alpha_i : \alpha \in \bar{A}, \pi^i(\alpha) = \beta\})$ and

$$\tilde{\alpha} = (\alpha_1, \dots, \alpha_{i-1}, 0, \alpha_{i+1}, \dots, \alpha_n).$$

(2.2) Then use the Add Pattern Routine 6.3 for

$$\tilde{\alpha} + \text{CH}(g\mathbf{e}^i, \frac{s}{g}) = \{(\alpha_1, \dots, \alpha_{i-1}, jg, \alpha_{i+1}, \dots, \alpha_n) : j = 0, \dots, \frac{s}{g}\}$$

to update \mathcal{P} and \bar{A} .

(3) Return the family of patterns \mathcal{P} and \bar{A} .

Observe that the output depends on the order of the enumeration of $[n]$ in step (2). Consider for example $\bar{A} = \{(1, 1), (1, 3), (2, 2)(4, 2)\}$. Analogously to the discussion after the Find Chain Pattern Routine, in step (2.2) there is a choice between the larger chain $\tilde{\alpha} + \text{CH}(\mathbf{e}^i, s)$ and the chain $\tilde{\alpha} + \text{CH}(g\mathbf{e}^i, \frac{s}{g})$. As before, we prefer to use smaller chains.

6.7 Find Axis Chain Pattern Routine

In order to find axis chains we modify the Find Shifted Chain Pattern Routine 6.6 by changing step (2) as follows: let $\beta = \mathbf{0}$ and, for all $i \in [n]$, which satisfy $\text{supp}(\alpha) = \{i\}$ for at least one $\alpha \in \bar{A}$, execute steps (2.1) and (2.2).

7 Computational Results

7.1 Setup

For a given finite and nonempty set $A \subseteq \mathbb{N}^n$ and a vector $\mathbf{f} \in \mathbb{R}^A$, the width function $\omega_{M^A(K)}(\mathbf{f})$ can be expressed as

$$\omega_{M^A(K)}(\mathbf{f}) = \max_{x \in K} f(x) - \min_{x \in K} f(x).$$

Thus, determination of the width function requires solving two instances of (POP). Choosing a family \mathcal{P} of patterns, we bound $\omega_{M^A(K)}(\mathbf{f})$ from above by

relaxing both (POP) to two instances of (P-RLX) with the objective functions $\langle -\mathbf{f}, \mathbf{v} \rangle$ and $\langle \mathbf{f}, \mathbf{v} \rangle$, respectively. The upper bound obtained this way is denoted by $\omega(\mathcal{P}, M^A(K), \mathbf{f})$. The values $\omega_{M^A(K)}(\mathbf{f})$ and $\omega(\mathcal{P}, M^A(K), \mathbf{f})$ depend on the length of the coefficient vector \mathbf{f} , and so we rescale them appropriately in such a way that the quality of the upper bound $\omega(\mathcal{P}, M^A(K), \mathbf{f})$ on the value $\omega_{M^A(K)}(\mathbf{f})$ is measured relatively to the quality of the (trivial) relaxation with just singleton patterns. More precisely, consider the family $\mathcal{P}_A^{\text{single}} = \{\{\alpha\} : \alpha \in A\}$ the family of all singleton patterns for A . The value $\omega(\mathcal{P}_A^{\text{single}}, M^A(K), \mathbf{f})$ can be determined explicitly, as mentioned in Subsection 4.1.

Thus, we compare

$$\nu(\mathcal{P}, A, \mathbf{f}) := \frac{\omega(\mathcal{P}, M^A(K), \mathbf{f})}{\omega(\mathcal{P}_A^{\text{single}}, M^A(K), \mathbf{f})}.$$

against

$$\nu_A(\mathbf{f}) := \frac{\omega_{M^A(K)}(\mathbf{f})}{\omega(\mathcal{P}_A^{\text{single}}, M^A(K), \mathbf{f})}.$$

We use the Cutting-Plane Algorithm 6.1 to compute $\omega(\mathcal{P}, M^A(K), \mathbf{f})$ and Baron to approximate $\omega_{M^A(K)}(\mathbf{f})$. Within the CPU time limit of 1000 seconds, Baron either returns a numerical approximation of $\omega_{M^A(K)}(\mathbf{f})$ or otherwise terminates providing a so-called best feasible intermediate solution. An approximation of the value $\nu_A(\mathbf{f})$ determined by Baron in this way is called the reference solution in what follows. We ran Baron 1.8.2 with standard settings on a 64-bit processor with 16 cores under LINUX Debian 3.2.

A set of test of instances consists of a nonempty and finite set $A \subseteq \mathbb{N}^n$ and 100 randomly sampled coefficient vectors $\mathbf{f}_1, \dots, \mathbf{f}_{100} \in \mathbb{R}^A$ from the uniform distribution in $[-1, 1]^A$. For all instances we choose $K = [0, 1]^n$. As A we use the sets A_1, \dots, A_6 from Figure 2, dense sets \mathbb{N}_d^3 for different d and sets A_7, \dots, A_9 , that are given by Figure 11. The distributions of $\nu(\mathcal{P}, A, \mathbf{f}_1), \dots, \nu(\mathcal{P}, A, \mathbf{f}_{100})$ for various families of patterns \mathcal{P} are visualized in box plots in Figures 9, 10 and 12.

Our implementation of the Cutting-Plane Algorithm 6.1 uses the following specifications. We choose $\varepsilon = 10^{-4}$. The separation problem for chains and shifted chains was solved using their LP approximations given by Corollary 5 and Corollary 8. For the covering \mathcal{I} we always choose an interval decomposition of 9 intervals of the same length. In our implementation of the pattern generation routine, the pattern types are enumerated in the following order:

1. Multilinear Patterns (ML)
2. Axis Chain Patterns (AC)
3. Chain Patterns (CH)
4. Shifted Chain Patterns (SC).

Axis chains are only used together with multilinear patterns.

7.2 Example Sets A_1, \dots, A_6

The sets A_1, \dots, A_6 from Figure 2 are chosen to illustrate the strengths and weaknesses of different pattern types and their combinations. If one pattern includes all elements of a set A , the relaxation (P-RLX) is tight. This is for example the case in Figure 4, the complete monomial structure of A_2 is captured by the chain

$$\text{CH}((1, 1), 5) = \{(0, 0), (1, 1), (2, 2), (3, 3), (4, 4), (5, 5)\}.$$

Hence the moment body $M^{A_2}([0, 1]^2)$ is perfectly approximated by the pattern relaxation using just the pattern $\text{CH}((1, 1), 5)$. This is reflected by Figure 9 as the box plot for the chains matches the box plot of the reference solution for the instances given for A_2 . The same holds for multilinear patterns and A_1 (Figure 3) as well as shifted chains and A_4 (Figure 5). The family of multilinear

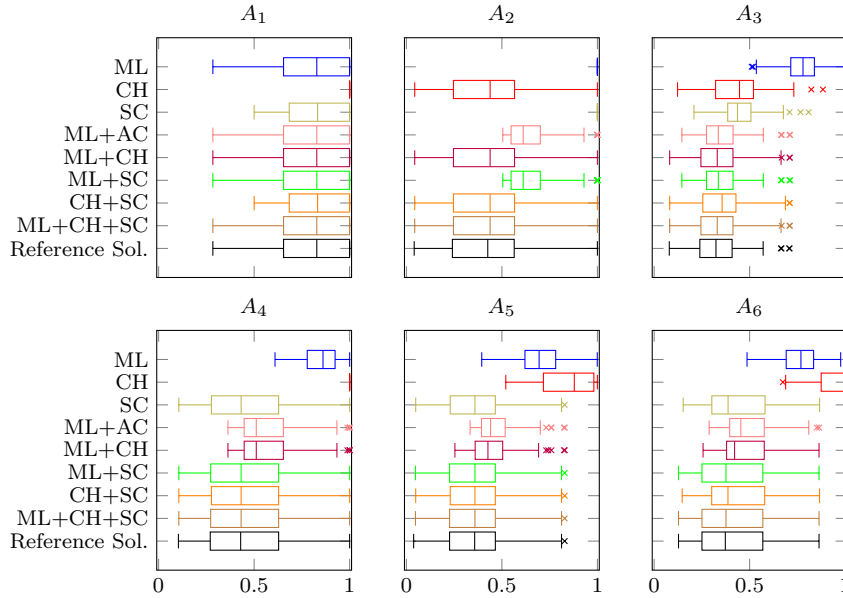


Fig. 9 The box plots visualize the distribution of the upper bounds $\nu_{A_i}^{\mathcal{P}}(\mathbf{f}_1), \dots, \nu_{A_i}^{\mathcal{P}}(\mathbf{f}_{100}), i \in [6]$, for different families of \mathcal{P} corresponding to the labels of the y axis. The box borders are 1/4 and 3/4-quantiles. The lower whisker is the smallest data value which is larger than lower quartile -1.5 times the interquartile range and the upper whisker is the largest data value which is smaller than upper quartile $+1.5$ times the interquartile range.

patterns $\mathcal{P}_{A_2}^{\text{ML}}$ in Figure 3 for the set A_2 does not connect any exponents in A_2 . As a consequence we have

$$\nu(\mathcal{P}_{A_2}^{\text{ML}}, A_2, \mathbf{f}_i) = 1$$

for all $i \in [100]$ (see Figure 9). The bounds on the width of $M^{A_2}(K)$ obtained by $\mathcal{P}_{A_2}^{\text{ML}}$ and the trivial relaxation coincide

$$\omega(\mathcal{P}_{A_2}^{\text{ML}}, M^A(K), \mathbf{f}_i) = \omega(\mathcal{P}_{A_2}^{\text{single}}, M^A(K), \mathbf{f}).$$

Surprisingly, Baron fails to obtain the optimal solution in 9% of all calculated widths for A_2 within the given time.

Even though some elements of A_5 or A_6 are only indirectly connected by shifted chains, Figure 9 suggests, that they yield reasonably tight relaxations of the moment bodies $M^{A_5}([0, 1]^2)$ and $M^{A_6}([0, 1]^2)$. For example, the two shifted chains

$$\begin{aligned} (3, 0) + \text{CH}((0, 1), 3) &= \{(3, 0), (3, 1), (3, 2), (3, 3)\}, \\ (4, 0) + \text{CH}((0, 1), 4) &= \{(4, 0), (4, 1), (4, 2), (4, 3), (4, 4)\} \end{aligned}$$

are enough to cover all elements of A_5 . The remaining shifted chains depicted in Figure 5 ensure that very moment variable \mathbf{v}_α with $\alpha \in (3, 0) + \text{CH}((0, 1), 3)$ is at least indirectly connected to every moment variable \mathbf{v}_β with $\beta \in (4, 0) + \text{CH}((0, 1), 4)$.

For the set A_3 , the family of multilinear patterns $\mathcal{P}_{A_3}^{\text{ML}}$ depicted in Figure 3 covers the entire set. Since the elements of different patterns in $\mathcal{P}_{A_3}^{\text{ML}}$ are not connected, this configuration does not perform well for the relaxation of $M^{A_3}([0, 1]^2)$. The same holds for chains and A_3 . By combining multilinear pattern with chains we are able to connect all element of A_3 at least indirectly. This combination improves the performance drastically compared to using just one of these pattern types. Even combining just the two axis chains with the multilinear patterns yields much better lower bounds. Generally, the combination of multilinear patterns and axis chains (ML + AC) seems to work well even in the case where multilinear patterns or chains alone yield weak relaxations.

7.3 Dense Exponent Sets

Testing on dense sets $A = \mathbb{N}_d^3$ of degree d gives us further insights into the potential of the pattern approach. For dense sets, the enumeration order of the patterns in the Add Pattern Routine 6.3 has no impact on the output.

In Figure 10 we observe that for increasing d multilinear patterns seem to work well for the approximation of $\omega_{M^{\mathbb{N}_d^3}(K)}(\mathbf{f})$. It seems that the family of multilinear patterns $\{ML(\alpha) : \alpha \in \mathbb{N}_d^3\}$ for the dense set \mathbb{N}_d^3 has nice connectivity properties. This might be the theoretical reason of our experimental observation. On the other hand, since families of chain patterns do not connect moment variables corresponding from different chains, using just chains alone results in weak lower bounds for $\omega_{M^{\mathbb{N}_d^3}(K)}(\mathbf{f})$.

Shifted chains with generators $\gamma = \mathbf{e}^i$ exploit parallel structures within A , whereas chains exploit diagonal structures. Combining the two for the dense

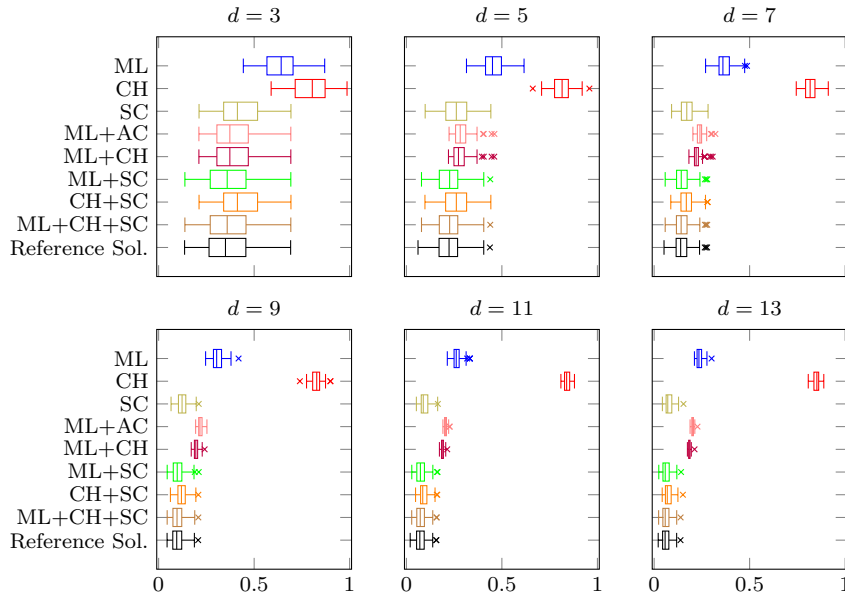


Fig. 10 The box plots visualize the distribution of the upper bounds $\nu_{\mathbb{N}_d^3}^{\mathcal{P}}(\mathbf{f}_1), \dots, \nu_{A_i}^{\mathcal{P}}(\mathbf{f}_{100})$, $d = 3, 5, 7, 9, 11, 13$, for different families of \mathcal{P} corresponding to the labels of the y axis. Box plot setting as in Figure 9.

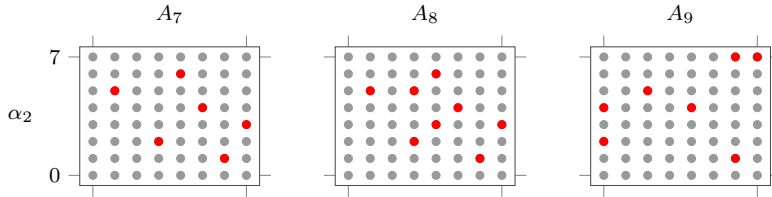


Fig. 11 The sets A_7, \dots, A_9 , depicted in red, have little structure to exploit by patterns.

set \mathbb{N}_d^3 yields results that are comparable to the reference solution. We also observe that by combining all pattern types we obtain solutions that are even close to the reference solutions. It would be interesting to find a theoretical justification for this.

7.4 Random sparse sets

The sets A_1, \dots, A_6 and dense sets \mathbb{N}_d^3 have structures that can obviously be exploited by our pattern types. In contrast to those sets, we consider sparse A_7, \dots, A_9 from Figure 11 that do not exhibit any particular structure. Using just one pattern type for the relaxation of instances given by sets A_7, \dots, A_9 yields similar lower bounds to the ones obtained using the trivial relaxation with just singleton patterns. As one can see from Figure 12, combinations of

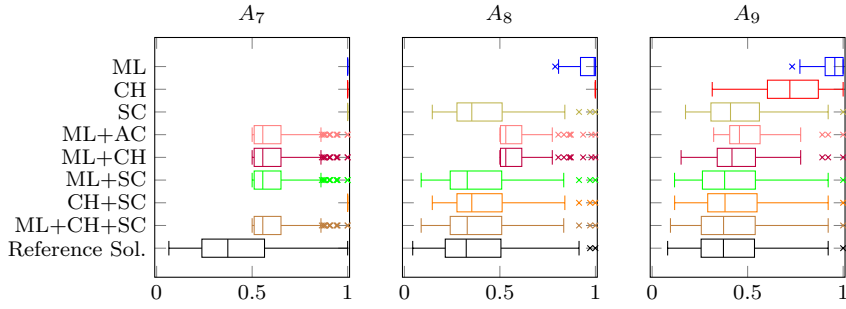


Fig. 12 The box plots visualize the distribution of the upper bounds $\nu_{A_i}^{\mathcal{P}}(\mathbf{f}_1), \dots, \nu_{A_i}^{\mathcal{P}}(\mathbf{f}_{100}), i = 7, 8, 9$, for different families of \mathcal{P} corresponding to the labels of the y axis. Box plot setting as in Figure 9.

several pattern types generate reasonable tight bounds in these cases. Moreover, for A_8 and A_9 , the combination of all tested patterns almost matches the reference solution.

8 Conclusion

We have presented a novel approach for the relaxation of polynomial optimization problems over a box that is based on patterns. The main advantage of our approach is that by using patterns we gain flexibility in terms of the size of the relaxation. The computational results suggest that we are able to generate reasonably tight lower bounds from the pattern relaxations. It therefore seems worthwhile to exploit the combinatorial structure of the set A of monomial exponents. Using the structure of A , we are able to neglect dependencies between certain monomials and avoid hard problem formulations and instead focus on well-behaved and easy-to-describe dependencies between certain other monomials. In doing so we produce tractable and sufficiently tight relaxations of (POP).

The Cutting-Plane Algorithm 6.1 is customizable and can be used with more involved patterns such as k -variate truncated submonoid patterns. We believe that by choosing an appropriate set of generators of a truncated submonoid pattern, the Cutting-Plane Algorithm 6.1 provides a way to make sum-of-squares methods applicable to polynomial problems of higher degree and more variables. In the same way, other Positivstellensätze such as SONC [14], can be incorporated into this framework.

Our cuts generated by (SP) can be integrated directly into divide-and-conquer frameworks that use moment variables, like BARON [29], SCIP [30], COUENNE [8] or LINDOGlobal [26]. The separation problem (3) can also be used as an interface to combine sum-of-squares methods with divide-and-conquer frameworks.

Acknowledgements The authors gratefully acknowledge support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - 314838170, GRK 2297 Math-CoRe. The second author was partially supported by the Ministry of Economy, Science and Digitalisation of the federal state Saxony-Anhalt by means of the Landesgraduierten-Stipendium.

References

1. Adjiman, C., Androulakis, I., Floudas, C.: A global optimization method, α BB, for general twice-differentiable constrained NLPs – II. Implementation and computational results. *Computers & Chemical Engineering* **22**(9), 1159–1179 (1998). DOI [http://dx.doi.org/10.1016/S0098-1354\(98\)00218-X](http://dx.doi.org/10.1016/S0098-1354(98)00218-X). URL <http://www.sciencedirect.com/science/article/pii/S009813549800218X>
2. Adjiman, C., Floudas, C.: Rigorous convex underestimators for general twice-differentiable problems. *Journal of Global Optimization* **9**, 23–40 (1996). DOI [10.1007/BF00121749](https://doi.org/10.1007/BF00121749). URL <http://dx.doi.org/10.1007/BF00121749>
3. Ahmadi, A.A., Majumdar, A.: DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. arXiv preprint arXiv:1706.02586 (2017)
4. Androulakis, I., Maranas, C., Floudas, C.: α BB: A global optimization method for general constrained nonconvex problems. *Journal of Global Optimization* **7**(4), 337–363 (1995). DOI [10.1007/BF01099647](https://doi.org/10.1007/BF01099647). URL <http://dx.doi.org/10.1007/BF01099647>
5. Anjos, M.F., Lasserre, J.B. (eds.): *Handbook on semidefinite, conic and polynomial optimization, International Series in Operations Research & Management Science*, vol. 166. Springer, New York (2012). DOI [10.1007/978-1-4614-0769-0](https://doi.org/10.1007/978-1-4614-0769-0). URL <https://doi.org/10.1007/978-1-4614-0769-0>
6. Averkov, G.: Constructive proofs of some Positivstellensätze for compact semialgebraic subsets of \mathbb{R}^d . *J. Optim. Theory Appl.* **158**(2), 410–418 (2013). DOI [10.1007/s10957-012-0261-9](https://doi.org/10.1007/s10957-012-0261-9). URL <https://doi.org/10.1007/s10957-012-0261-9>
7. Averkov, G.: Optimal size of linear matrix inequalities in semidefinite approaches to polynomial optimization. arXiv preprint arXiv:1806.08656 (2018)
8. Belotti, P.: Couenne, an exact solver for nonconvex minlps (2015). URL <https://projects.coin-or.org/Couenne/>
9. Boland, N., Dey, S.S., Kalinowski, T., Molinaro, M., Rigterink, F.: Bounding the gap between the McCormick relaxation and the convex hull for bilinear functions. *Mathematical Programming* **162**(1-2), 523–535 (2017)
10. Boyd, S., Vandenberghe, L.: *Convex optimization*. Cambridge University Press, Cambridge (2004). DOI [10.1017/CBO9780511804441](https://doi.org/10.1017/CBO9780511804441). URL <https://doi.org/10.1017/CBO9780511804441>
11. Chandrasekaran, V., Shah, P.: Relative entropy relaxations for signomial optimization. *SIAM J. Optim.* **26**(2), 1147–1173 (2016). DOI [10.1137/140988978](https://doi.org/10.1137/140988978). URL <https://doi.org/10.1137/140988978>
12. Chandrasekaran, V., Shah, P.: Relative entropy optimization and its applications. *Math. Program.* **161**(1-2, Ser. A), 1–32 (2017). DOI [10.1007/s10107-016-0998-2](https://doi.org/10.1007/s10107-016-0998-2). URL <https://doi.org/10.1007/s10107-016-0998-2>
13. Dalkiran, E., Sherali, H.D.: Theoretical filtering of RLT bound-factor constraints for solving polynomial programming problems to global optimality. *J. Global Optim.* **57**(4), 1147–1172 (2013). DOI [10.1007/s10898-012-0024-z](https://doi.org/10.1007/s10898-012-0024-z). URL <https://doi.org/10.1007/s10898-012-0024-z>
14. Dressler, M.: Sums of nonnegative circuit polynomials. Ph.D. thesis, Goethe-Universität Frankfurt am Main (2018)
15. Dressler, M., Ilman, S., de Wolff, T.: A Positivstellensatz for sums of nonnegative circuit polynomials. *SIAM J. Appl. Algebra Geom.* **1**(1), 536–555 (2017). DOI [10.1137/16M1086303](https://doi.org/10.1137/16M1086303). URL <https://doi.org/10.1137/16M1086303>
16. Dressler, M., Naumann, H., Theobald, T.: The dual cone of sums of non-negative circuit polynomials. arXiv preprint arXiv:1809.07648 (2018)

17. Duffin, R.J., Peterson, E.L., Zener, C.: Geometric programming: Theory and application. John Wiley & Sons, Inc., New York-London-Sydney (1967)
18. Ghasemi, M., Marshall, M.: Lower bounds for a polynomial on a basic closed semialgebraic set using geometric programming. arXiv preprint arXiv:1311.3726 (2013)
19. Iliman, S., de Wolff, T.: Amoebas, nonnegative polynomials and sums of squares supported on circuits. *Res. Math. Sci.* **3**, Paper No. 9, 35 (2016). DOI 10.1186/s40687-016-0052-2. URL <https://doi.org/10.1186/s40687-016-0052-2>
20. Lasserre, J.B.: An introduction to polynomial and semi-algebraic optimization. Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge (2015). DOI 10.1017/CBO9781107447226. URL <https://doi.org/10.1017/CBO9781107447226>
21. Laurent, M.: Sums of squares, moment matrices and optimization over polynomials. In: Emerging applications of algebraic geometry, *IMA Vol. Math. Appl.*, vol. 149, pp. 157–270. Springer, New York (2009). DOI 10.1007/978-0-387-09686-5_7. URL https://doi.org/10.1007/978-0-387-09686-5_7
22. Marshall, M.: Positive polynomials and sums of squares, *Mathematical Surveys and Monographs*, vol. 146. American Mathematical Society, Providence, RI (2008). DOI 10.1090/surv/146. URL <https://doi.org/10.1090/surv/146>
23. Mitsos, A., Chachuat, B., Barton, P.: McCormick-Based Relaxations of Algorithms. *SIAM Journal on Optimization* **20**(2), 573–601 (2009). DOI 10.1137/080717341. URL <http://dx.doi.org/10.1137/080717341>
24. Putinar, M.: Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.* **42**(3), 969–984 (1993). DOI 10.1512/iumj.1993.42.42045. URL <https://doi.org/10.1512/iumj.1993.42.42045>
25. Schneider, R.: Convex bodies: the Brunn-Minkowski theory, *Encyclopedia of Mathematics and its Applications*, vol. 151, expanded edn. Cambridge University Press, Cambridge (2014)
26. Schrage, L.E.: Optimization modeling with LINGO. Lindo System (2006)
27. Seidler, H., de Wolff, T.: An experimental comparison of sonc and sos certificates for unconstrained optimization. arXiv preprint arXiv:1808.08431 (2018)
28. Smith, E., Pantelides, C., Reklaitis, G.: A symbolic reformulation spatial branch-and-bound algorithm for the global optimization of nonconvex minlp's. *Computers & Chemical Engineering* **25**, 1399–1401 (2001). DOI 10.1016/S0098-1354(01)00733-5
29. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. *Math. Program.* **103**(2), 225–249 (2005)
30. Vigerske, S., Gleixner, A.: SCIP: global optimization of mixed-integer nonlinear programs in a branch-and-cut framework. *Optim Methods Softw* **33**(3), 563–593 (2018). DOI 10.1080/10556788.2017.1335312. URL <https://doi.org/10.1080/10556788.2017.1335312>