Monomial Patterns in Polynomial Optimization

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium

(Dr. rer. nat.)

von M.Sc. Benjamin Peters

geb. am 12.07.1989 in Burg

genehmigt durch die Fakultät für Mathematik

der Otto–von–Guericke–Universität Magdeburg

Gutachter: Prof. Dr. Gennadiy Averkov

Prof. Dr. Sebastian Sager

Prof. Dr. Christoph Helmberg

eingereicht am: 25.01.2021

Verteidigung am: 09.04.2021

Benjamin Peters

M.Sc.

Institut für Mathematische Optimierung

Otto–von–Guericke–Universität Magdeburg

Gebäude 02

Universitätsplatz 2

39106 Magdeburg

Germany

benjamin.peters@ovgu.de

Abstract

Convexification is a core technique in global polynomial optimization, which is used to generate convex relaxations of a polynomial optimization problem (POP). These relaxations in turn allow to compute bounds on the optimal value of a POP.

Currently, there are two main convexification approaches competing in theory and practice: the approach of nonlinear programming and the approach based on positivity certificates from real algebra. The former is comparatively cheap from a computational point of view, but typically does not provide tight relaxations with respect to bounds for the original problem. The latter is typically computationally expensive, but provides tight relaxations.

We embed both kinds of approaches into a unified framework of monomial relaxations. This framework of monomial relaxations is based on groups of exponents, which we call patterns. In order to build a relaxation, the POP is linearized by replacing each of its monomials with a monomial variable. Then the monomial variables that are indexed by the exponents of a pattern are linked by convex constraints. By identifying the appropriate patterns and their associated constraints, a variety of established convexification methods can be expressed within this framework. These include convexification methods based on sum-of-squares polynomials or nonnegative circuit polynomials as as well as multilinear envelopes. Within our framework we can freely combine the different patterns and their constraints. The combination of the different patterns allows to exploit the monomial structure of the polynomial problem. By selecting appropriate combinations of patterns we can trade off the quality of the bounds against computational expenses. Thus, it is possible to develop custom-made convexification strategies that are fitted to the problem structure and the demands of the user. Examples of different such strategies are given.

Furthermore, we develop a new pattern type called truncated submonoid and determine the corresponding convex constraints.

Different relaxations that are derived from combinations of patterns are numerically tested on self-generated benchmark instances. The computational experiments yield very encouraging results.

Zusammenfassung

Konvexifizierung ist eine Kerntechnik der globalen polynomiellen Optimierung, um konvexe Relaxierungen von polynomiellen Optimierungsproblemen (POP) zu erstellen. Diese Relaxierungen ermöglichen es wiederum, Schranken für das Minimum eines POP zu berechnen.

Derzeit gibt es zwei Hauptansätze, die in Theorie und Praxis miteinander konkurrieren: den Ansatz der nichtlinearen Optimierung und einen Ansatz, der auf Positivitätszertifikaten aus der reellen Algebra basiert. Ersterer ist aus rechnerischer Sicht vergleichsweise günstig, bietet jedoch typischerweise schwache Relaxierungen, was die berechneten Schranken betrifft. Letzterer ist typischerweise rechenintensiv, liefert jedoch starke Relaxierungen.

Wir betten beide Ansätze in einen einheitlichen Rahmen von monomiellen Relaxierungen ein. Diese monomiellen Relaxierungen basieren auf Mengen von Exponenten, die wir Patterns nennen. Um eine Relaxierung zu erstellen, wird das POP linearisiert, indem Monome durch Monomenvariablen ersetzt werden. Dann werden die Monomenvariablen, die durch die Exponenten eines Patterns indiziert sind, durch konvexe Nebenbedingungen miteinander verknüpft. Durch Identifizieren geeigneter Patterns und der damit verbundenen Nebenbedingungen kann eine Vielzahl etablierter Konvexifizierungsmethoden auf diese Weise ausgedrückt werden. Dazu gehören Methoden, die auf Quadraten von Polynomen oder nichtnegativen Circuit-Polynomen basieren, sowie multilineare Relaxierungen.

Innerhalb unserer Relaxierung können wir die verschiedenen Patterns und deren Nebenbedingungen frei kombinieren. Die Kombination verschiedener Patterns ermöglicht es, die monomielle Struktur eines POP auszunutzen. Mithilfe geeigneter Kombinationen kann die Qualität der Schranken gegen den Rechenaufwand abgewägt werden. Somit ist es möglich, maßgeschneiderte Konvexifizierungsstrategien zu entwickeln, die an die Problemstruktur und die Anforderungen des Benutzers angepasst sind. Wir geben Beispiele verschiedener solcher Strategien.

Darüber hinaus entwickeln wir einen neuen Patterntyp namens Truncated Submonoid und bestimmen für diesen die entsprechenden konvexen Nebenbedingungen.

Außerdem werden verschiedene, durch Patterns induzierte, Relaxierungen anhand eigener Benchmark-Instanzen numerisch getestet. Die Computerexperimente liefern vielversprechende Ergebnisse.

Contents

Chapter 1

Introduction

In this thesis, we provide a template for the computation of lower bounds for polynomial optimization problems (POPs). The problems that we consider are of the following general type: Let $\mathrm{n} \in \mathbb{N}$ with $\mathrm{n} > 0$ and $\mathrm{r} \in \mathbb{N}$. Given a polynomial objective function $f := \sum_{\alpha \in \mathbb{N}^{\mathrm{n}}} \mathrm{f}_\alpha x^\alpha \in \mathbb{R}[x]$ and polynomial constraints $g^1, \dots, g^{\mathrm{r}} \in \mathbb{R}[x]$ in n indeterminates $x := (x_1, \dots, x_{\mathrm{n}})$, determine the infimum f over the set $\mathrm{K} := \{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^1(\mathbf{x}) \geq 0, \dots, g^{\mathrm{r}}(\mathbf{x}) \geq 0\}$:

$$\text{minimize} \quad f(\mathbf{x}) \quad \text{subject to} \quad \mathbf{x} \in \mathrm{K}. \tag{POP}$$

Here, $x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_{\mathrm{n}}^{\alpha_{\mathrm{n}}}$ and f_α denote the monomial and the coefficient corresponding to the exponent α, respectively. This problem type includes, among others, linear, quadratic and 0-1 programming. It comes therefore to no surprise that hard problems can be encoded within this format. In fact it is known that minimizing a multivariate polynomial of degree as little as 4 over $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$ is in general $\mathcal{NP}$-hard [44]. This prompts the necessity for lower bounding schemes that allow to tackle (POP).

POPs appear in a variety of different areas such as in discrete and combinatorial optimization, control systems and robotics, statistics and electric power systems engineering, to name a few. For a discussion of specific problems from these areas through the lens of POPs see for example [1, Sec. 1.1], [38, Sec. 1.1] and [14, Sec. 3.6].

Furthermore, POPs also arise as intermediate problems in global optimization algorithms, which utilize symbolic reformulation [59, 62]. There, factorable functions of a global optimization problem (GOP) are replaced by polynomial functions. This way one obtains an intermediate POP, whose optimal value is a lower bound on the optimal value of the original GOP. Usually, the intermediate POP is further relaxed in order to actually compute a lower bound. We illustrate this procedure by a small example.

Example 1.1.
One possible way[1] *to construct a relaxation of the problem*

$$\text{minimize} \quad \sin(\mathrm{x})^2 e^{\mathrm{x}} + \sin(\mathrm{x}) \quad \text{subject to} \quad \mathrm{x} \in [0, \tfrac{\pi}{2}]$$

is to replace $\sin(\mathrm{x})$ *and* e^{x} *with auxiliary variables* y_1 *and* y_2*, respectively. The corresponding intermediate POP is*

$$\text{minimize} \quad \mathrm{y}_1^2\mathrm{y}_2 + \mathrm{y}_1 \quad \text{subject to} \quad \mathrm{y}_1 \in [0,1], \mathrm{y}_2 \in [1, e^{\frac{\pi}{2}}]. \tag{1.1}$$

Note that the lower/upper bounds of y_1 *and* y_2 *are the minimum/maximum of* $\sin(\mathrm{x})$ *and* e^{x} *over* $\mathrm{x} \in [0, \frac{\pi}{2}]$*. Replacing* y_1^2 *with* y_3 *yields the multilinear problem*

$$\text{minimize} \quad \mathrm{y}_1\mathrm{y}_3\mathrm{y}_2 + \mathrm{y}_1 \quad \text{subject to} \quad \mathrm{y}_1 \in [0,1], \mathrm{y}_2 \in [1, e^{\frac{\pi}{2}}], \mathrm{y}_3 \in [0,1].$$

For the latter problem exist convexification strategies based on multilinear envelopes [43, 15]. △

Many lower bounds obtained from the relaxations of such intermediate POPs are utilized within a spatial branch-and-bound framework to approximate an optimal solution of the original GOP. The intermediate POPs pose especially intriguing applications of (POP) as one can tackle problems like (1.1) directly with methods from polynomial optimization. Thus, they constitute an interface between the 'global optimization community' and the 'polynomial optimization community'. In fact, they show how the development of new relaxation methods for POPs may advance an, until now, almost entirely separate branch of global optimization methods. In turn polynomial optimization approaches can benefit from the branching schemes that have been developed and refined in the last two decades. Integrating polynomial optimization methods within popular global optimization solvers, like `BARON`, may also help to promote solution methods that are otherwise tailored to POPs.[2]

What makes (POP) interesting for optimization from a methodological standpoint, is that, despite its generality, it still possesses structures that make it accessible to optimization algorithms. In order to find and exploit these structure we will resort to looking at (POP) from the perspective of moments and from the perspective of

[1]It might be that a particular solver handles this example differently.
[2]In particular `BARON` is a widely distributed global optimization solver. On the website of `BARON` https://minlp.com/about it is stated that '[o]ver 600 organizations now license `BARON` in more than 50 countries'.

nonnegative polynomials. The moment perspective of (POP) is

$$\text{minimize} \sum_{\alpha \in \mathbb{N}^n} \mathrm{f}_\alpha \mathrm{v}_\alpha \ \text{ subject to } \ (\mathrm{v}_\alpha)_{\alpha \in \mathbb{N}^n} \text{ is a moment sequence of a probability measure on } \mathrm{K}.$$

It is known that the objective value of the above coincides with the one of (POP) [38, Ch. 4.2]. The question whether $(\mathrm{v}_\alpha)_{\alpha \in \mathbb{N}^n}$ is such a moment sequence is commonly referred to as K-moment problem [37, Ch. 2.7]. The dual perspective of the above is

$$\text{maximize } \lambda \text{ subject to } \lambda \in \mathbb{R} \text{ and } f(\mathbf{x}) - \lambda \geq 0 \text{ for all } \mathbf{x} \in \mathrm{K}.$$

Apparently, this problem also has the same objective value as (POP). Here, the underlying question is whether $f - \lambda$ is nonnegative on K.

Both questions have been subject to lively research in the past. There, the relationship of moments and nonnegative polynomials with sum-of-squares (SOS) polynomials has been of particular interest. In 1927 Hilbert's 17th Problem, that is the question if all on $\mathbb{R}^n$ nonnegative polynomials can be represented as a sum of squares of rational functions, was solved by Artin [5]. After that a variety of nonnegativity certificates based on SOS polynomials were discovered by Krivine [34], Stengle [64] or Putinar [51], just to name a few. The fact that positive semidefinite (PSD) matrices can be used to represent SOS polynomials and the emergence of efficient semidefinite programming (SDP) solvers has brought these certificates into the focus of the optimization community. The certificates combined with PSD representations of SOS polynomials lead to convex SDP relaxations of (POP) that can be solved by SDP solvers. The driving forces behind the development of different, mostly hierarchical, SOS/SDP approaches for the computation of lower bounds, were, among others, Shor [60], Nesterov [45], Parrilo [48] and Lasserre [35]. Besides SOS/PSD based certificates there exist certificates based on geometric and linear programming, such as a sum of nonnegative circuit polynomials (SONC) Positivstellensatz [29] and Handelman's hierarchy [27].

In general many important convexification techniques applied to POPs share the following common distinctive features: monomials x^α are substituted with monomial variables v_α and the relationships among them are captured, exactly or in a relaxed 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 'global optimization community' uses monomial variables and constraints such that the resulting relaxations are cheap to solve. The resulting poor lower bounds are compensated by solving many relaxations within a branch-and-bound framework. The 'polynomial optimization community' usually aims to solve only one single relaxation, which often produces a very tight bound. This comes at the price of a large number of monomial variables and hard constraints. Interestingly, so far there has been little interaction between the two different schools of thought. We believe that a major reason is the lack of a mathematical formalism that would allow a uniform description of different convexification techniques.

One contribution of this thesis is the introduction of the notion of patterns to fill this gap. Patterns are finite sets $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ of exponent vectors. It is our goal to determine patterns P such that the monomial variables v_α indexed by $\alpha \in \mathrm{P}$ can be linked by constraints that satisfy a given demand on the computability. One may express this goal also in terms of nonnegative polynomials: determine P such that there exists a relaxation of the cone

$$\mathcal{P}(\mathrm{K})_{\mathrm{P}} := \Big\{ p = \sum_{\alpha \in \mathrm{P}} \mathrm{p}_\alpha x^\alpha : p(\mathbf{x}) \geq 0 \ \text{ for all } \ \mathbf{x} \in \mathrm{K} \Big\},$$

that satisfies a given demand on the computability and tightness. While various kinds of patterns have been implicitly used by the disjoint research communities, the introduction of the explicit notion of patterns allows to develop a unifying mathematical language that highlights common ideas. Promoting this elementary notion enables to see similarities of the different research directions and will help to connect different communities that work independently on the same problems. For example, the pattern

$$\mathrm{P} = \{(1,0),(0,1),(1,1)\}$$

corresponds to the well-known McCormick envelope [43, 15], i.e. the convexification of the variables x_1 and x_2 and their product $\mathrm{x}_1\mathrm{x}_2$ over an axis-parallel box K. Other examples of methods that can be expressed using the notion of patterns are moment relaxations and their dual the sum-of-squares relaxations [2, 38, 40], scaled-diagonally-dominant sums of squares (SDSOS) [1], SONC [21, 58], bound-factor products [19] and their dual the Handelman's hierarchy [27], multilinear intermediates [11], polyhedral outer approximations [66] as well as expression trees [62, 59].

Furthermore, we propose the pattern relaxation, that is a flexible template for the relaxation of POPs. Within this template we can combine different types of patterns to build convex relaxations of (POP). This allows us to use the ideas of these, until now, largely disjoint schools of thought together. Our new and more general point of view might also help to understand numerical issues and the facial structures of feasible sets in the aforementioned convexification approaches. This, in turn, can be expected to have a positive impact on the improvement of existing and on the development of novel approaches to polynomial optimization. Apart from expressing the already established convexification methods as patterns, we derive various new convexification techniques from the pattern template. The resulting relaxations can be solved by a variety of different numerical approaches.

In the interest of analyzing the tightness and computational expenses related to different convexification strategies, we use a self-implemented, customizable prototype cutting-plane algorithm and, as an alternative, an implementation that employs the interior point solver `MOSEK` [3]. While the cutting-plane algorithm is not necessarily the best practice or a novel approach, the choice is a reasonable standard strategy for sparse convex problems. Furthermore, the warm start capabilities of linear programming solvers as well as the particular chosen implementation make this algorithm highly customizable and demonstrate how new pattern types can be used for strengthening already existing relaxations. On the other hand, the usage of the interior point solver allows to benefit from the state-of-the-art algorithms and barrier functions implemented in `MOSEK`.

1.1 Thesis Overview

Parts of this thesis are published in [9] and its general structure follows [9]. Corollaries, lemmas, propositions and theorems that can be found in the current version of [9] are marked.

The thesis is organized as follows. The basic notation is given in **Section 1.2**. In **Chapter 2** we first introduce the notion of the pattern relaxation from the moment perspective and then deduce a characterization in terms of nonnegative polynomials. Furthermore, the separation problem for patterns is formulated as an optimization problem, which is later used in the cutting-plane algorithm. **Chapter 3** is dedicated to the interpretation and discussion of established convexification techniques as monomial patterns. This also includes a generalization of multilinear envelopes as multilinear patterns and characterizations of the SONC cone and SDSOS cone derived from their respective patterns. In particular, the there given characterization of SONC cone allows to combine it with other pattern types within the pattern relaxation. In **Chapter 4** a new pattern type is introduced, which generalizes the patterns corresponding to moment relaxations, SOS and SDSOS relaxations. This pattern type gives rise to a variety of sparse relaxations for unconstrained as well as constrained problems. **Chapter 5** is dedicated to the algorithms used for the computations. There, a prototype of a hybrid cutting-plane algorithm for pattern relaxations is presented and an implementation that utilizes a conic programming solver is discussed. Additionally, a prototype routine for the systematic selection of patterns is presented on the example of multilinear patterns. Computational results in **Chapter 6** highlight the benefits of our novel approach. Finally, the conclusion and outlook are given in **Chapter 7**.

1.2 Notation

In this section we set up the notation, terminologies and conventions that are needed to smoothly read this thesis. These conventions hold unless stated otherwise. This part is complemented by the back matter starting from page 111.

The number n is a positive integer and the numbers d, r are nonnegative integers. $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ is always a nonempty and finite set whose elements usually are multi-exponents or multi-indices that are contextually linked to exponents. The set $\{1, \ldots, \mathrm{n}\}$ is denoted by $[\mathrm{n}]$ and $[\mathrm{n}]_0 := [\mathrm{n}] \cup \{0\}$. The cardinality of a set A is denoted by $\#\mathrm{A}$. We heavily rely on indexing using sets like A. For example, the vector space of real vectors which are indexed by A is

$$\mathbb{R}^{\mathrm{A}} := \{(\mathrm{v}_\alpha)_{\alpha \in \mathrm{A}} : \mathrm{v}_\alpha \in \mathbb{R} \text{ for all } \alpha \in \mathrm{A}\}.$$

For a vector $\mathbf{v} = (\mathrm{v}_\alpha)_{\alpha \in \mathrm{A}}$ of $\mathbb{R}^{\mathrm{A}}$ we use boldface and upshape print and for its components only upshape. While we portray a vector $\mathbf{v}$ for convenience as $1 \times \mathrm{A}$ row vector, we interpret it in matrix multiplications as $\mathrm{A} \times 1$ column vector. The ℓ_1-norm and the ℓ_∞-norm of $\mathbf{v}$ are denoted by $||\mathbf{v}||_1$ and $||\mathbf{v}||_\infty$. Note that by ordering A we can always identify $\mathbb{R}^{\mathrm{A}}$ with $\mathbb{R}^{\mathrm{k}}$ for $\mathrm{k} = \#\mathrm{A}$. The all-zero vector of $\mathbb{R}^{\mathrm{A}}$ is denoted by $\mathbf{0}^{\mathrm{A}}$, the all-one vector by $\mathbf{1}^{\mathrm{A}}$ and the standard basis vectors are denoted by $\mathbf{e}^{\alpha,\mathrm{A}}$, i.e.

$$\mathrm{e}^{\alpha,\mathrm{A}}_\beta := \begin{cases} 1 & \text{if } \beta = \alpha, \\ 0 & \text{else} \end{cases} \qquad \text{for all } \alpha, \beta \in \mathrm{A}.$$

When it causes no ambiguity, we omit the A and write $\mathbf{0}$, $\mathbf{1}$ and $\mathbf{e}^\alpha$ instead. $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ is a closed set that contains a full-dimensional (open) ball $\{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : ||\mathbf{x} - \mathbf{a}||_1 < \varepsilon\}$ for some $\varepsilon > 0$ and $\mathbf{a} \in \mathbb{R}^{\mathrm{n}}$. Since all the emerging spaces in this thesis are equipped with a standard ℓ-norm, we use as topology the standard topology induced by the respective norms. For the rest of this section let $\mathrm{B} \subseteq \mathbb{N}^{\mathrm{n}}$ be a nonempty set (possibly of infinite cardinality) and let $\mathbf{v} \in \mathbb{R}^{\mathrm{A}}$ and $\mathbf{w} \in \mathbb{R}^{\mathrm{B}}$ be two vectors. The support of $\mathbf{w}$ is

$$\mathrm{supp}(\mathbf{w}) := \{\beta \in \mathrm{B} : \mathrm{w}_\beta \neq 0\}.$$

If $\mathrm{A} \cap \mathrm{B} \neq \emptyset$, we define the bilinear product of $\mathbf{v}$ and $\mathbf{w}$ as

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

If $\mathrm{B} \subseteq \mathrm{A}$, we denote the coordinate projections onto components indexed by B of the vector $\mathbf{v}$ and a nonempty set $\mathrm{S} \subseteq \mathbb{R}^{\mathrm{A}}$ by

$$\mathbf{v}_{\mathrm{B}} := (\mathrm{v}_\alpha)_{\alpha \in \mathrm{B}} \qquad \text{and} \qquad [\![\mathrm{S}]\!]_{\mathrm{B}} := \{\mathbf{v}_{\mathrm{B}} : \mathbf{v} \in \mathrm{S}\}.$$

If $\mathrm{B} = \{\beta\}$ contains just a single element, we use v_β or $[\![\mathbf{v}]\!]_\beta$ and $[\![\mathrm{S}]\!]_\beta$ instead. Furthermore, cl(S) is the topological closure of S, conv(S) the convex hull of S and cone(S) the convex conic hull of S. Throughout this thesis, we use 'cone' in place of 'convex cone'. Frequently used cones are: the nonnegative real numbers $\mathbb{R}_+$, the (strictly) positive real numbers $\mathbb{R}_{++}$ and the cone of positive semidefinite (PSD) $\mathrm{A} \times \mathrm{A}$ matrices $\mathbb{S}^{\mathrm{A}}_+$. The dual of a cone $\mathcal{K}$ is denoted by $\mathcal{K}^*$. Let $x = (x_1, \ldots, x_\mathrm{n})$ be a vector of n indeterminates and $\alpha \in \mathbb{N}^\mathrm{n}$ a multi-exponent. We call

$$x^\alpha := x_1^{\alpha_1} \cdot \ldots \cdot x_\mathrm{n}^{\alpha_\mathrm{n}}$$

a monomial and refer to its evaluation $\mathbf{x}^\alpha$ at a point $\mathbf{x} \in \mathbb{R}^\mathrm{n}$ also as monomial. The minimum and maximum of $\mathbf{x}^\alpha$ over a compact set $\mathrm{K} \subseteq \mathbb{R}^\mathrm{n}$ are

$$\underline{\mathbf{x}}^\alpha_\mathrm{K} := \min_{\mathbf{x} \in \mathrm{K}} \mathbf{x}^\alpha \quad \text{and} \quad \overline{\mathbf{x}}^\alpha_\mathrm{K} := \max_{\mathbf{x} \in \mathrm{K}} \mathbf{x}^\alpha,$$

respectively, $\underline{\mathbf{x}}^\mathrm{A}_\mathrm{K} := (\underline{\mathbf{x}}^\alpha_\mathrm{K})_{\alpha \in \mathrm{A}}$ and $\overline{\mathbf{x}}^\mathrm{A}_\mathrm{K} := (\overline{\mathbf{x}}^\alpha_\mathrm{K})_{\alpha \in \mathrm{A}}$. As usual, $\mathbb{R}[x]$ is the ring of polynomials

$$f = \sum_{\alpha \in \mathbb{N}^\mathrm{n}} \mathrm{f}_\alpha x^\alpha,$$

where $\mathbf{f} = (\mathrm{f}_\alpha)_{\alpha \in \mathbb{N}^\mathrm{n}} \in \mathbb{R}^{\mathbb{N}^\mathrm{n}}$ is a coefficient vector with finite support, i.e. $\#\,\mathrm{supp}(\mathbf{f}) < \infty$. By slight abuse of notation we define the monomial support of a polynomial f as

$$\mathrm{supp}(f) := \mathrm{supp}(\mathbf{f}).$$

We obtain the coefficients of f that are indexed by a nonempty set $\mathrm{B} \subseteq \mathbb{N}^\mathrm{n}$ using

$$\mathrm{vec}(f)_\mathrm{B} := \mathbf{f}_\mathrm{B}.$$

A polynomial is usually denoted by a letter in italic print and, when no ambiguity can arise, we use the same letter in bold face print to denote its coefficient vector, that is $\mathbf{f} \in \mathbb{R}^\mathrm{B}$ for an appropriate set B with $\mathrm{supp}(f) \subseteq \mathrm{B}$. If we want to emphasize the underlying vector x of intermediates, we write $f(x)$ instead of f. The degree of a finite set B is defined as

$$\deg(\mathrm{B}) := \max\{|\beta_1| + \cdots + |\beta_\mathrm{n}| : \beta \in \mathrm{B}\}$$

and the degree of a polynomial $f \in \mathbb{R}[x]$ as $\deg(f) := \deg(\mathrm{supp}(f))$. By

$$\mathcal{P}(\mathrm{K}) := \{f \in \mathbb{R}[x] : f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \in \mathrm{K}\}.$$

we denote the cone of polynomials that are nonnegative over K. Throughout this thesis we mostly work in the truncated version of $\mathbb{R}[x]$. That is by A we prescribe which monomials x^α can occur in f with nonzero coefficient f_α:

$$\begin{aligned}\mathbb{R}[x]_{\mathrm{A}} &:= \Big\{f \in \mathbb{R}[x] : \mathrm{supp}(f) \subseteq \mathrm{A}\Big\} \\ &= \Big\{f \in \mathbb{R}[x] : f = \sum_{\alpha \in \mathrm{A}} \mathrm{f}_\alpha x^\alpha \text{ with } \mathbf{f} \in \mathbb{R}^{\mathrm{A}}\Big\}.\end{aligned}$$

By means of the coefficient vectors we can identify $\mathbb{R}[x]_{\mathrm{A}}$ with $\mathbb{R}^{\mathrm{A}}$. Analogously, we define

$$\mathcal{P}(\mathrm{K})_{\mathrm{A}} := \Big\{f \in \mathcal{P}(\mathrm{K}) : \mathrm{supp}(f) \subseteq \mathrm{A}\Big\}.$$

The norm of a polynomial $f \in \mathbb{R}[x]_{\mathrm{A}}$ is defined as $||f||_1 := ||\mathbf{f}_{\mathrm{A}}||_1$. Note that $\mathcal{P}(\mathrm{K})_{\mathrm{A}}$ is a closed and convex cone. The A-truncated moment vector map or simply moment vector is

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

Since it will come in handy later on, we also allow sets $\mathrm{Z} \subseteq \mathbb{Z}^{\mathrm{n}}$ as index sets for $\mathrm{m}(x)_{\mathrm{Z}}$. In order to avoid confusing the i-th power of a polynomial p with a superscript, we write $(p)^{\mathrm{i}}$. For example $(p^{\mathrm{j}})^{\mathrm{i}}$ is the i-th power of the polynomial p^{j}. A polynomial $p \in \mathbb{R}[x]$ is called SOS if it is the sum of finitely many squared polynomials, that is $p = (p^1)^2 + \cdots + (p^{\mathrm{k}})^2$ for polynomials $p^1, \ldots, p^{\mathrm{k}} \in \mathbb{R}[x]$. The cone of n-variate SOS with monomial support in B is

$$\mathrm{SOS}(\mathrm{B}) := \Big\{p \in \mathbb{R}[x]_{\mathrm{B}} : p \text{ is SOS}\Big\}.$$

Clearly, it always holds $\mathrm{SOS}(\mathrm{B}) \subseteq \mathcal{P}(\mathrm{K})_{\mathrm{B}}$. A (basic closed) semialgebraic set is a subset of $\mathbb{R}^{\mathrm{n}}$ defined by polynomial inequalities

$$\Big\{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^1(\mathbf{x}) \geq 0, \ldots, g^{\mathrm{r}}(\mathbf{x}) \geq 0\Big\},$$

where $g^1, \ldots, g^{\mathrm{r}} \in \mathbb{R}[x]$ are polynomials. Greek letters, with a few exceptions, are reserved for multi-exponents or multi-indices that are related to multi-exponents. Let $\mathrm{X} \subseteq \mathbb{R}^{\mathrm{A}}$ and $c : \mathrm{X} \longrightarrow \mathbb{R}$ be a continuous function. We write the optimization problem of minimizing the objective function c over the feasible set X as

$$\begin{aligned} &\text{minimize} && c(\mathbf{v}) \\ &\text{for} && \mathbf{v} \in \mathbb{R}^{\mathrm{A}} \\ &\text{subject to} && \mathbf{v} \in \mathrm{X}. \end{aligned} \tag{OP}$$

The optimal value $\mathrm{c}^{\mathrm{op}} \in \mathbb{R} \cup \{-\infty, \infty\}$ of the minimization problem (OP) is defined as

$$\mathrm{c}^{\mathrm{op}} := \inf\{c(\mathbf{v}) : \mathbf{v} \in \mathrm{X}\}.$$

We say that $\mathbf{v} \in \mathbb{R}^{\mathrm{A}}$ is a feasible solution for (OP) if $\mathbf{v} \in \mathrm{X}$ and refer to a feasible point $\mathbf{v}^\star$ with $c(\mathbf{v}^\star) = \mathrm{c}^{\mathrm{op}}$ as an optimal solution of (OP). If an optimal solution exists, we say that the optimal value is attained. The existence of an optimal solution immediately implies that c^{op} is finite. (OP) is said to be infeasible if $\mathrm{X} = \emptyset$ and otherwise feasible. If $\mathrm{c}^{\mathrm{op}} = -\infty$, (OP) is said to be unbounded and if $\mathrm{c}^{\mathrm{op}} > -\infty$, it is called bounded. By replacing 'minimize' with 'maximize', 'inf' with 'sup' and '$\mathrm{c}^{\mathrm{op}} > -\infty$' with '$\mathrm{c}^{\mathrm{op}} < \infty$' we acquire the respective definitions for maximization problems.

Chapter 2

Pattern Relaxation

In the first part of this chapter we develop the theory of patterns and pattern relaxations from the perspective of monomial variables. By dualizing the pattern relaxation in the second part we obtain a characterization in terms of nonnegative polynomials. Both views will pave the way to incorporate well established relaxation techniques into our framework. At last the separation problem for patterns is formulated and we show how new pattern types can be created by shifting already existing types.

2.1 Monomial Convexification and Monomial Relaxation

We consider the problem of minimizing a polynomial $f \in \mathbb{R}[x]_{\mathrm{A}}$ over the set K:

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

Via lifting, we reformulate (POP) as an optimization problem over a manifold in $\mathbb{R}^{\mathrm{A}}$ with a linear objective:

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

In (S-POP) we have replaced each monomial $\mathbf{x}^\alpha$ by a variable v_α. We therefore call the components v_α of $\mathbf{v}$ in (S-POP) monomial variables. For the next step we introduce the following definition:

Definition 2.1 (Truncated Moment Body and Truncated Moment Cone). *The* (A*-truncated) moment body is defined as*

$$\mathcal{M}(\mathrm{K})_{\mathrm{A}} := \operatorname{cl}\operatorname{conv}\left(\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\}\right)$$

and the (A*-truncated) moment cone as*

$$\mathcal{C}(\mathrm{K})_{\mathrm{A}} := \operatorname{cl}\operatorname{cone}(\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\}). \qquad \triangle$$

Replacing the feasible set $\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\}$ of (S-POP) by its convex hull $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ yields the monomial convexification of (POP):

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

The convexification (C-POP) of (POP) is tight, that is, the optimal values of (C-POP) and (POP) coincide. Unfortunately for general sets A, the constraint $\mathbf{v} \in \mathcal{M}(\mathrm{K})_{\mathrm{A}}$ is difficult to verify. Thus, it is natural to relax $\mathbf{v} \in \mathcal{M}(\mathrm{K})_{\mathrm{A}}$ to a system of simpler constraints of the same type.

Definition 2.2 (Pattern, Pattern Family and Pattern Relaxation). *We call a finite and nonempty set $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ pattern and a finite and nonempty family of patterns $\mathcal{F}$ pattern family. If the pattern family $\mathcal{F}$ satisfies*

$$\mathrm{A} \subseteq \bigcup_{\mathrm{P} \in \mathcal{F}} \mathrm{P}, \tag{2.1}$$

we refer to $\mathcal{F}$ as pattern relaxation of A*. The corresponding system*

$$\mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}} \ \text{ for } \ \mathrm{P} \in \mathcal{F} \tag{2.2}$$

is called pattern relaxation of $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ (with respect to $\mathcal{F}$) and

$$\mathbf{v}_{\mathrm{P}} \in \mathcal{C}(\mathrm{K})_{\mathrm{P}} \ \text{ for } \ \mathrm{P} \in \mathcal{F} \tag{2.3}$$

is called (conic) pattern relaxation of $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ (with respect to $\mathcal{F}$). Throughout, we use $\mathrm{A}_{\mathcal{F}}$ to denote $\bigcup_{\mathrm{P} \in \mathcal{F}} \mathrm{P}$. $\triangle$

It is our intention to cover A by patterns P such that the corresponding moment bodies $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$ yield more structure that we can exploit algorithmically than the original moment body $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$. Using a pattern family $\mathcal{F}$ that satisfies (2.1) we lift (C-POP), that is

$$\begin{array}{cl} \text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} \\ \text{subject to} & \mathbf{v} \in \mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}, \end{array} \tag{C-POP'}$$

and then relax the latter by

$$\begin{array}{cl} \text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} \\ \text{subject to} & \mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}} \ \text{ for all } \ \mathrm{P} \in \mathcal{F}. \end{array} \tag{P-RLX'}$$

The optimal values of (C-POP) and of (C-POP') coincide.[1] Hence, the optimal value $\mathrm{f}^{\mathrm{PRLX}'}$ of (P-RLX') is a lower bound on the optimal value $\mathrm{f}^{\mathrm{pop}}$ of (POP). We refer to (P-RLX') as pattern relaxation of (POP) as it does not conflict with Definition 2.2. By relaxing the moment bodies in (P-RLX') with their respective moment cones we obtain the conic version of the pattern relaxation:

$$\begin{array}{cl} \text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} \\ \text{subject to} & \mathbf{v}_{\mathrm{P}} \in \mathcal{C}(\mathrm{K})_{\mathrm{P}} \ \text{ for all } \ \mathrm{P} \in \mathcal{F} \\ & \mathrm{v}_{\mathbf{0}} = 1. \end{array} \tag{P-RLX}$$

Here, we implicitly assume that $\mathrm{A}_{\mathcal{F}}$ contains $\mathbf{0}$.[2] Since for every feasible solution $\mathbf{v}$ of (P-RLX') it holds $\mathrm{v}_{\mathbf{0}} = 1$, the optimal value $\mathrm{f}^{\mathrm{PRLX}}$ of (P-RLX) satisfies $\mathrm{f}^{\mathrm{PRLX}} \leq \mathrm{f}^{\mathrm{PRLX}'}$.

The following lemma shows that (P-RLX) and (P-RLX') agree if $\mathbf{0} \in \mathrm{P}$ for all $\mathrm{P} \in \mathcal{F}$. Hence, we will make no linguistic differentiation between (P-RLX) and (P-RLX') and also refer to (P-RLX) as a pattern relaxation of (POP).

[1] Recall that the bilinear product in (C-POP') boils down to $\langle \mathbf{f}, \mathbf{v} \rangle = \sum_{\alpha \in \mathrm{A}} \mathrm{f}_\alpha \mathrm{v}_\alpha$ since $\mathrm{A} \subseteq \mathrm{A}_{\mathcal{F}}$.
[2] Since $\mathcal{M}(\mathrm{K})_{\{\mathbf{0}\}} = \{1\}$, we can always add the pattern $\mathrm{P} = \{\mathbf{0}\}$ to $\mathcal{F}$ if $\mathbf{0} \notin \mathrm{A}_{\mathcal{F}}$.

Lemma 2.3.
For a pattern $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ *with* $\mathbf{0} \in \mathrm{P}$ *holds* $\mathcal{C}(\mathrm{K})_{\mathrm{P}} \cap \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{P}} : \mathrm{v}_{\mathbf{0}} = 1\right\} = \mathcal{M}(\mathrm{K})_{\mathrm{P}}$. △

Proof. Observe that $\mathcal{C}(\mathrm{K})_{\mathrm{P}} = \operatorname{cl}\operatorname{cone}(\mathcal{M}(\mathrm{K})_{\mathrm{P}})$. Thus,

$$\begin{aligned}&\mathcal{C}(\mathrm{K})_{\mathrm{P}} \cap \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{P}} : \mathrm{v}_{\mathbf{0}} = 1\right\} = \\ &\operatorname{cl}\left\{\lambda\mathbf{w} \in \mathbb{R}^{\mathrm{P}} : \mathbf{w} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}}, \lambda \geq 0 \text{ with } \lambda\mathbf{w}_{\mathbf{0}} = 1\right\}.\end{aligned} \tag{2.4}$$

From $\mathbf{x}^{\mathbf{0}} = 1$ for all $\mathbf{x} \in \mathrm{K}$ follows $\mathrm{w}_{\mathbf{0}} = 1$ for all $\mathbf{w} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}}$. Hence, $\lambda\mathrm{w}_{\mathbf{0}} = 1$ in (2.4) implies $\lambda = 1$. This shows that the right hand side of (2.4) is identical to $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$. □

Lemma 2.3 shows that (P-RLX) is equivalent to

$$\begin{array}{lll}\text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle & \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} & \\ \text{subject to} & \mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}} & \text{for all } \mathrm{P} \in \mathcal{F} \text{ with } \mathbf{0} \in \mathrm{P} \\ & \mathbf{v}_{\mathrm{P}} \in \mathcal{C}(\mathrm{K})_{\mathrm{P}} & \text{for all } \mathrm{P} \in \mathcal{F} \text{ with } \mathbf{0} \notin \mathrm{P}.\end{array}$$

Observe that when K is compact, so is $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$. If additionally to that there exists for each $\alpha \in \mathrm{A}$ a pattern $\mathrm{P} \in \mathcal{F}$ with $\alpha, \mathbf{0} \in \mathrm{P}$, then the optimal value of (P-RLX) is guaranteed to be bounded.

This procedure of developing the pattern relaxation can also be seen as embedding $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ into $\mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}$ for some set $\mathrm{A}_{\mathcal{F}}$ that contains A and can be represented nicely as a union of patterns $\mathrm{P} \in \mathcal{F}$. Geometrically, the passage from (POP) through (C-POP) and (C-POP') to (P-RLX') can be represented by the diagram

$$\mathrm{m}(\mathrm{K})_{\mathrm{A}} \xrightarrow{\text{convexifying}} \mathcal{M}(\mathrm{K})_{\mathrm{A}} \xrightarrow{\text{embedding}} \mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}} \xrightarrow{\text{projecting}} \mathcal{M}(\mathrm{K})_{\mathrm{P}}. \tag{2.5}$$

A visualization of this diagram can be found in Figure 2.1. There, the procedure is also depicted for the set $\mathrm{A} = \{1, 2, 3\}$, $\mathrm{K} = [0, 1]$ and the pattern family $\mathcal{F} = \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. Note that in this example we do not have an actual embedding, since $\mathrm{A}_{\mathcal{F}} = \mathrm{A}$. Furthermore, the example is a fairly simple one and the pattern relaxation should not be expected to perform as well as Figure 2.1 suggests.

An advantage of the pattern approach is that we can choose patterns $\mathrm{P} \in \mathcal{F}$ such that the computational costs to solve (P-RLX') or (P-RLX) and quality of the obtained lower bounds on the objective value of (POP) are well balanced. In view of this, we are interested in the choice of patterns P, for which the computational costs of enforcing the constraints $\mathbf{v}_\mathrm{P} \in \mathcal{M}(\mathrm{K})_\mathrm{P}$ or $\mathbf{v}_\mathrm{P} \in \mathcal{C}(\mathrm{K})_\mathrm{P}$ meet our requirements. Since (2.1) is an inclusion and not an equality, we can find such patterns even if A is ill-structured. This is the reason for introducing auxiliary moment variables v_β, $\beta \in \mathrm{A}_\mathcal{F} \backslash \mathrm{A}$ in (C-POP'), which are not present in (C-POP).

The quality of a pattern relaxation $\mathcal{F}$ depends on how the moment variables are connected by the system of conditions (2.2). We present a simple definition of a connectivity notion that will later on help to explain and compare different pattern families.

Definition 2.4 (Directly and indirectly connected).
We say that monomial variables v_α and v_β or the exponents α and β are directly connected by a pattern family $\mathcal{F}$ if $\alpha, \beta \in \mathrm{P}$ holds for some $\mathrm{P} \in \mathcal{F}$. Furthermore, v_α and v_β or α and β are (indirectly) connected by $\mathcal{F}$ if there exist a sequence of patterns $\{\mathrm{P}^\mathrm{j}\}_{\mathrm{j}\in[\mathrm{k}]} \in \mathcal{F}$ such that $\alpha \in \mathrm{P}^1, \beta \in \mathrm{P}^\mathrm{k}$ and $\mathrm{P}^\mathrm{j} \cap \mathrm{P}^{\mathrm{j}+1} \setminus \{\mathbf{0}\} \neq \emptyset$ for all $\mathrm{j} \in [\mathrm{k}-1]$. △

We will see in Chapters 3 and 4 that there exist various pattern types. An open question is how to best select a pattern family from this multitude of patterns to construct a pattern relaxation. The next lemma shows that certain patterns dominate others and therefore allows to avoid including unnecessary combinations of patterns within a pattern family.

Lemma 2.5.
Let P and $\tilde{\mathrm{P}}$ be two patterns with $\mathrm{P} \subseteq \tilde{\mathrm{P}}$. Then $\mathbf{v} \in \mathcal{C}(\mathrm{K})_{\tilde{\mathrm{P}}}$ implies $\mathbf{v}_\mathrm{P} \in \mathcal{C}(\mathrm{K})_\mathrm{P}$. △

Proof. Since $\mathbf{v} \in \mathcal{C}(\mathrm{K})_{\tilde{\mathrm{P}}}$, there exists a sequence $\{\mathbf{v}^{(\mathrm{k})}\}_{\mathrm{k}\in\mathbb{N}} \subseteq \mathcal{C}(\mathrm{K})_{\tilde{\mathrm{P}}}$ that satisfies

- for each k there exist $\lambda^{(\mathrm{k},\mathrm{i})} \in \mathbb{R}_+$ and $\mathbf{x}^{(\mathrm{k},\mathrm{i})} \in \mathrm{K}$ with $\mathrm{i} \in [\mathrm{r}]$ and $\mathrm{r} \leq \#\tilde{\mathrm{P}}+1$ such that $\mathbf{v}^{(\mathrm{k})} = \sum_{\mathrm{i}\in[\mathrm{r}]} \lambda^{(\mathrm{k},\mathrm{i})}\, \mathrm{m}(\mathbf{x}^{(\mathrm{k},\mathrm{i})})_{\tilde{\mathrm{P}}}$
- $\lim_{\mathrm{k}\to\infty} \mathbf{v}^{(\mathrm{k})} = \mathbf{v}$.

Defining $\mathbf{w}^{(\mathrm{k})} := \sum_{\mathrm{i}\in[\mathrm{r}]} \lambda^{(\mathrm{k},\mathrm{i})}\, \mathrm{m}(\mathbf{x}^{(\mathrm{k},\mathrm{i})})_\mathrm{P}$, we see that $\{\mathbf{w}^{(\mathrm{k})}\}_{\mathrm{k}\in\mathbb{N}} \subseteq \mathcal{C}(\mathrm{K})_\mathrm{P}$ and

$$\lim_{\mathrm{k}\to\infty} \mathbf{w}^{(\mathrm{k})} = \mathbf{v}_\mathrm{P} \in \mathcal{C}(\mathrm{K})_\mathrm{P}.$$

□

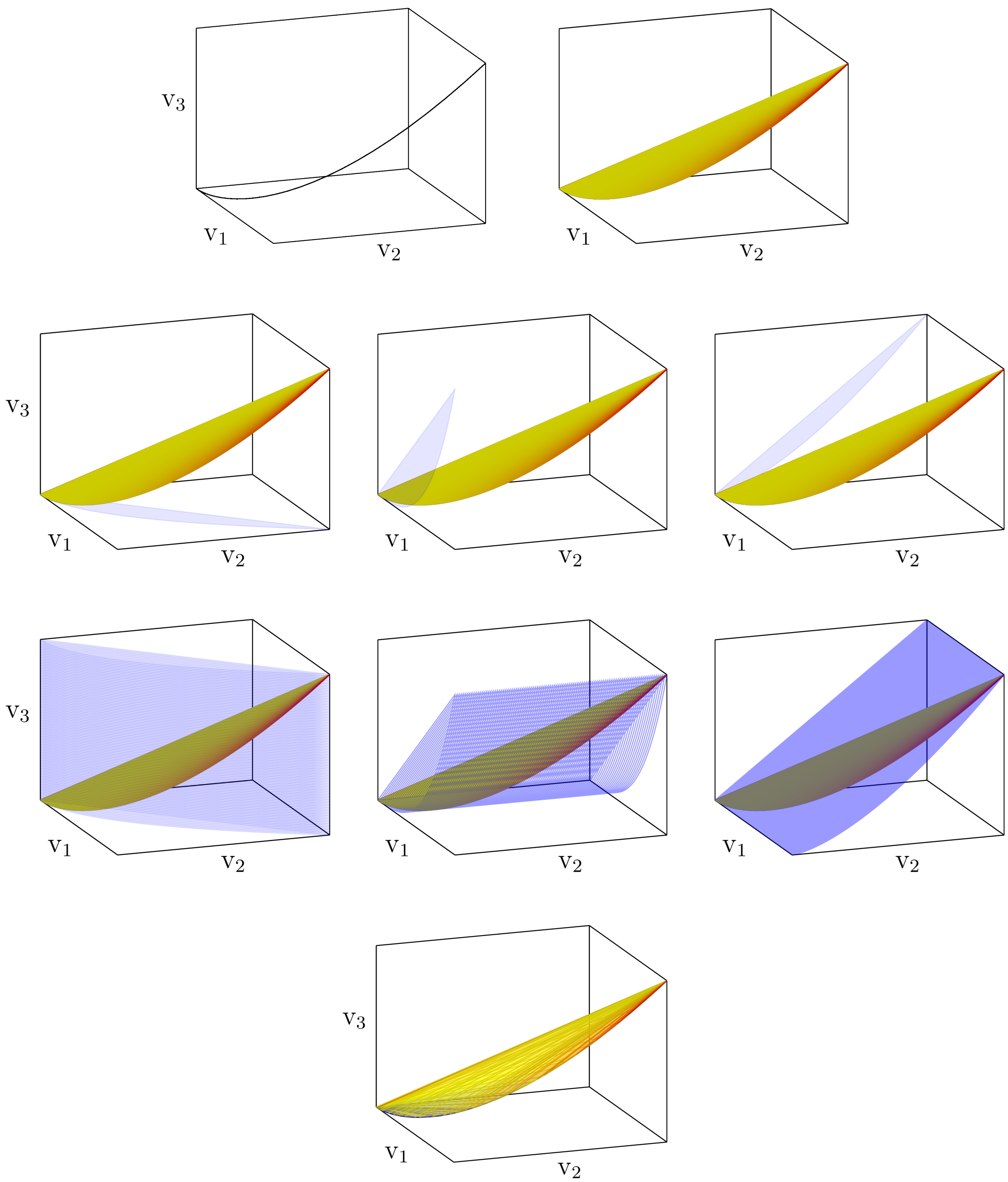

Figure 2.1: **First row left:** the curve $\{m(\mathbf{x})_A : \mathbf{x} \in K\}$ for $A = \{1, 2, 3\}$ and $K = [0, 1]$. **First row right:** the corresponding moment body $\mathcal{M}(K)_A$. **Second row from left to right:** the projections $\mathcal{M}(K)_P$ of $\mathcal{M}(K)_A$ in blue for $P \in \mathcal{F}$ with $\mathcal{F} := \{\{1, 2\}, \{1, 3\}, \{2, 3\}\}$. **Third row from left to right:** the liftings $\mathbb{R}^{A \setminus P} \times \mathcal{M}(K)_P$ of $\mathcal{M}(K)_P$ for $P \in \mathcal{F}$. **Last row:** the pattern relaxation $\{\mathbf{v} : \mathbf{v}_P \in \mathcal{M}(K)_P \text{ for } P \in \mathcal{F}\}$ of $\mathcal{M}(K)_A$ that is induced by the pattern family $\mathcal{F}$.

2.2 Dual of (P-RLX)

In this part we derive the dual of the pattern relaxation from (P-RLX). It is well known that untruncated moment cones and cones of positive polynomials with untruncated monomial support are dual to each other, see for example [38, Sec. 1.2]. We briefly prove that this also holds for truncated moment cones.

Lemma 2.6.
The dual cone $\mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}$ of $\mathcal{P}(\mathrm{K})_{\mathrm{A}}$ is $\mathcal{C}(\mathrm{K})_{\mathrm{A}}$. △

Proof. From the representation of the cone of nonnegative polynomials $\mathcal{P}(\mathrm{K})_{\mathrm{A}}$

$$\mathcal{P}(\mathrm{K})_{\mathrm{A}} = \left\{ f \in \mathbb{R}[x]_{\mathrm{A}} : \langle \mathbf{f}, \mathrm{m}(\mathbf{x})_{\mathrm{A}} \rangle \geq 0 \ \text{ for all } \ \mathbf{x} \in \mathrm{K} \right\} \tag{2.6}$$

and the definition of its dual cone

$$\mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*} = \left\{ \mathbf{v} \in \mathbb{R}^{\mathrm{A}} : \langle \mathbf{f}, \mathbf{v} \rangle \geq 0 \ \text{ for all } \ f \in \mathcal{P}(\mathrm{K})_{\mathrm{A}} \right\}$$

it follows that $\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} \in \mathbb{R}^{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\} \subseteq \mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}$. Since $\mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}$ is a closed and convex cone, we also have

$$\mathcal{C}(\mathrm{K})_{\mathrm{A}} = \operatorname{cl}\operatorname{cone}\left(\left\{ \mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K} \right\}\right) \subseteq \mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}. \tag{2.7}$$

For the reverse inclusion assume that there exists $\mathbf{v} \in \mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*} \backslash \mathcal{C}(\mathrm{K})_{\mathrm{A}}$. Then, by a separation theorem for closed convex cones, like [25, Thm. 4.4.2], there exists a vector $\mathbf{p} \in \mathbb{R}^{\mathrm{A}}$ such that $\langle \mathbf{p}, \mathbf{v} \rangle < 0$ and $\langle \mathbf{p}, \mathbf{w} \rangle \geq 0$ for all $\mathbf{w} \in \mathcal{C}(\mathrm{K})_{\mathrm{A}}$. In particular, $p = \langle \mathbf{p}, \mathrm{m}(x)_{\mathrm{A}} \rangle \in \mathcal{P}(\mathrm{K})_{\mathrm{A}}$, which is a contradiction to $\langle \operatorname{vec}(p)_{\mathrm{A}}, \mathbf{v} \rangle = \langle \mathbf{p}, \mathbf{v} \rangle < 0$ and $\mathbf{v} \in \mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}$. Hence, we have $\mathcal{C}(\mathrm{K})_{\mathrm{A}} = \mathcal{P}(\mathrm{K})_{\mathrm{A}}^{*}$. □

The next corollary is a standard fact from convex optimization and is proven for polar cones in [54, Cor. 16.4.2] .

Corollary 2.7.
Let $\mathcal{K}^{1}, \ldots, \mathcal{K}^{\mathrm{r}} \subseteq \mathbb{R}^{\mathrm{n}}$ be closed, convex and nonempty cones. Then

$$\Big(\sum_{\mathrm{i} \in [\mathrm{r}]} \mathcal{K}^{\mathrm{i}} \Big)^{*} = \bigcap_{\mathrm{i} \in [\mathrm{r}]} \left(\mathcal{K}^{\mathrm{i}} \right)^{*}.$$

△

Observe that $\mathcal{P}(\mathrm{K})_\mathrm{P}$ is an intersection of closed half-spaces, i.e.

$$\mathcal{P}(\mathrm{K})_\mathrm{P} = \bigcap_{\mathbf{x}\in \mathrm{K}} \left\{ f \in \mathbb{R}[x]_\mathrm{P} : \langle \mathbf{f}, \mathrm{m}(\mathbf{x})_\mathrm{P}\rangle \geq 0 \right\},$$

and therefore closed and convex.

Lemma 2.8.

Let $\mathcal{F}$ be a pattern family. Then

$$\left(\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}\right)^* = \bigcap_{\mathrm{P}\in\mathcal{F}} \mathcal{C}(\mathrm{K})_\mathrm{P} \times \mathbb{R}^{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}}.$$

△

Proof. All cones $\mathcal{P}(\mathrm{K})_\mathrm{P}$ are nonempty, as they contain the 0 polynomial. Furthermore, $\mathcal{P}(\mathrm{K})_\mathrm{P}$ is closed and convex for each $\mathrm{P} \in \mathcal{F}$. Hence, from Corollary 2.7 it follows

$$\left(\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}\right)^* = \bigcap_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}^*. \tag{2.8}$$

We interpret $\mathcal{P}(\mathrm{K})_\mathrm{P}$ as a cone that lives in $\mathbb{R}^\mathrm{P}$. That means we write (2.8) by abuse of notation as[3]

$$\left(\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P} \times \left\{\mathbf{0}^{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}}\right\}\right)^* = \bigcap_{\mathrm{P}\in\mathcal{F}} \left(\mathcal{P}(\mathrm{K})_\mathrm{P} \times \left\{\mathbf{0}^{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}}\right\}\right)^*. \tag{2.9}$$

From Lemma 2.6 it follows that

$$\left(\mathcal{P}(\mathrm{K})_\mathrm{P} \times \left\{\mathbf{0}^{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}}\right\}\right)^* = \mathcal{C}(\mathrm{K})_\mathrm{P} \times \mathbb{R}^{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}}$$

and therefore by (2.9) the assertion. □

Lemma 2.9.

Let $\mathcal{F}$ be a pattern family. Then

$$\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}$$

is a closed cone. △

[3]This representation emphasizes that for a polynomial $f^\mathrm{P} \in \mathcal{P}(\mathrm{K})_\mathrm{P}$ it holds $\mathrm{vec}(f^\mathrm{P})_{\mathrm{A}_\mathcal{F}\setminus \mathrm{P}} = \mathbf{0}$.

Proof. The cones $\mathcal{P}(\mathrm{K})_\mathrm{P}, \mathrm{P} \in \mathcal{F}$ are closed, convex and nonempty. Furthermore, $\sum_{\mathrm{P}\in\mathcal{F}} f^\mathrm{P} = 0$ with $f^\mathrm{P} \in \mathcal{P}(\mathrm{K})_\mathrm{P}$ for all $\mathrm{P} \in \mathcal{F}$ if and only if[4] $f^\mathrm{P} = 0$ for all $\mathrm{P} \in \mathcal{F}$. Thus, the assertion follows from [54, Cor. 9.1.3]. □

From Lemma 2.9 and Lemma 2.8 it immediately follows

$$\begin{aligned}\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P} &= \Big(\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}\Big)^{**} \\ &= \Big(\bigcap_{\mathrm{P}\in\mathcal{F}} \mathcal{C}(\mathrm{K})_\mathrm{P} \times \mathbb{R}^{\mathrm{A}_\mathcal{F}\setminus\mathrm{P}}\Big)^{*}.\end{aligned} \tag{2.10}$$

We can now dualize (P-RLX) under rather mild assumptions.

Theorem 2.10.
Let $\mathcal{F}$ be a pattern family such that $\mathrm{A} \subseteq \mathrm{A}_\mathcal{F}$ and $\mathbf{0} \in \mathrm{A}_\mathcal{F}$. For $f \in \mathbb{R}[x]_\mathrm{A}$ the objective value $\mathrm{f}^{\mathrm{PRLX}}$ of (P-RLX) *and the objective value $\mathrm{f}^{\mathrm{DRLX}}$ of*

$$\begin{aligned} &\text{maximize} && \lambda \\ &\text{for} && \lambda \in \mathbb{R} \\ &\text{subject to} && f - \lambda \in \sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P} \end{aligned} \tag{D-RLX}$$

coincide. △

Proof. For the second equality $(*)$ from below, we use $\mathbf{f} = \mathrm{vec}(f)_{\mathrm{A}_\mathcal{F}}$ and (2.10) in order to express the cone $\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}$ by its dual. The equality $(**)$ follows from Lemma 2.8. Hence,

$$\begin{aligned}\mathrm{f}^{\mathrm{DRLX}} &= \sup\Big\{\lambda \in \mathbb{R} : f - \lambda \in \sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}\Big\} \\ &\overset{(*)}{=} \sup\Big\{\lambda \in \mathbb{R} : \langle \mathbf{f} - \lambda \cdot \mathbf{e}^{\mathbf{0},\mathrm{A}_\mathcal{F}}, \mathbf{v}\rangle \geq 0 \text{ for all } \mathbf{v} \in \Big(\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_\mathrm{P}\Big)^{*}\Big\} \\ &\overset{(**)}{=} \sup\Big\{\lambda \in \mathbb{R} : \langle \mathbf{f}, \mathbf{v}\rangle \geq \lambda \cdot \mathrm{v}_\mathbf{0} \text{ for all } \mathbf{v} \in \bigcap_{\mathrm{P}\in\mathcal{F}} \mathcal{C}(\mathrm{K})_\mathrm{P} \times \mathbb{R}^{\mathrm{A}_\mathcal{F}\setminus\mathrm{P}}\Big\}.\end{aligned} \tag{2.11}$$

For a pattern P with $\mathbf{0} \in \mathrm{P}$ it follows from $\mathrm{m}(\mathbf{x})_\mathbf{0} = 1$ for all $\mathbf{x} \in \mathrm{K}$ that $\mathrm{w}_\mathbf{0} > 0$ for

[4] Recall that K contains a full-dimensional ball.

all $\mathbf{w}_{\mathrm{P}} \in \mathcal{C}(\mathrm{K})_{\mathrm{P}} \backslash \{\mathbf{0}\}$. Such a pattern P exists in $\mathcal{F}$ because $\mathbf{0} \in \mathrm{A}_{\mathcal{F}}$. Hence, $\mathrm{w}_{\mathbf{0}} > 0$ for all

$$\mathbf{w} \in \left\{ \mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} : \mathbf{v} \in \bigcap_{\mathrm{P} \in \mathcal{F}} \mathcal{C}(\mathrm{K})_{\mathrm{P}} \times \mathbb{R}^{\mathrm{A}_{\mathcal{F}} \backslash \mathrm{P}} \right\} \backslash \{\mathbf{0}\}. \tag{2.12}$$

Let $\mathrm{F} := \bigcap_{\mathrm{P} \in \mathcal{F}} \mathcal{C}(\mathrm{K})_{\mathrm{P}} \times \mathbb{R}^{\mathrm{A}_{\mathcal{F}} \backslash \mathrm{P}}$. Since $\mathrm{K} \neq \emptyset$, there exists $\mathbf{w} \in \mathrm{F} \backslash \{\mathbf{0}\}$. Observe that $\frac{1}{\mathrm{k}+1}\mathbf{w} \in \mathrm{F}$ for all $\mathrm{k} \in \mathbb{N}$ and $\lim_{\mathrm{k} \to \infty} \frac{1}{\mathrm{k}+1}\mathbf{w} = \mathbf{0}$. Thus, the set in (2.12) is dense in F. It follows from (2.11) that

$$\mathrm{f}^{\mathrm{DRLX}} = \sup \left\{ \lambda \in \mathbb{R} : \langle \mathbf{f}, \mathbf{v} \rangle \geq \lambda \ \text{ for all } \ \mathbf{v} \in \bigcap_{\mathrm{P} \in \mathcal{F}} \mathcal{C}(\mathrm{K})_{\mathrm{P}} \times \mathbb{R}^{\mathrm{A}_{\mathcal{F}} \backslash \mathrm{P}} \ \text{ with } \ \mathrm{v}_{\mathbf{0}} = 1 \right\}.$$

As the objective value of the above is strictly increasing in λ, we can compute $\mathrm{f}^{\mathrm{DRLX}}$ by

$$\begin{aligned} \mathrm{f}^{\mathrm{DRLX}} &= \inf \left\{ \langle \mathbf{f}, \mathbf{v} \rangle : \mathbf{v} \in \bigcap_{\mathrm{P} \in \mathcal{F}} \mathcal{C}(\mathrm{K})_{\mathrm{P}} \times \mathbb{R}^{\mathrm{A}_{\mathcal{F}} \backslash \mathrm{P}} \ \text{ and } \ \mathrm{v}_{\mathbf{0}} = 1 \right\} \\ &= \inf \left\{ \langle \mathbf{f}, \mathbf{v} \rangle : \mathbf{v}_{\mathrm{P}} \in \mathcal{C}(\mathrm{K})_{\mathrm{P}} \ \text{ for all } \ \mathrm{P} \in \mathcal{F} \ \text{ and } \ \mathrm{v}_{\mathbf{0}} = 1 \right\} \\ &= \mathrm{f}^{\mathrm{PRLX}}. \end{aligned}$$

□

Since $\mathrm{f}^{\mathrm{DRLX}}$ and the optimal value $\mathrm{f}^{\mathrm{PRLX}}$ of (P-RLX) coincide, we will also refer to (D-RLX) as pattern relaxation of (POP).

An alternative way of writing (D-RLX) is

$$\begin{aligned} \text{maximize} \quad & \lambda \\ \text{for} \quad & \lambda \in \mathbb{R}, \\ & f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}} \quad \text{for all } \mathrm{P} \in \mathcal{F} \\ \text{subject to} \quad & f - \lambda = \sum_{\mathrm{P} \in \mathcal{F}} f^{\mathrm{P}}. \end{aligned} \tag{2.13}$$

This formulation of (D-RLX) emphasises that the crucial optimization variables are the scalar λ and the polynomials f^{P} with restricted monomial support $\mathrm{supp}(f^{\mathrm{P}}) \subseteq \mathrm{P}$. We use superscript P to stress that a polynomial f^{P} corresponds to the pattern P.

Lemma 2.11.
If the objective value $\mathrm{f}^{\mathrm{DRLX}}$ *of* (2.13) *is finite, it is also attained.* △

Proof. That $\mathrm{f}^{\mathrm{DRLX}}$ is finite, implies that there exists a sequence $\{(f^{(\mathrm{P},\mathrm{i})})_{\mathrm{P}\in\mathcal{F}}, \lambda^{(\mathrm{i})}\}_{\mathrm{i}\in\mathbb{N}}$ of polynomials $f^{(\mathrm{P},\mathrm{i})} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}}$ for all $\mathrm{P} \in \mathcal{F}$ and scalars $\lambda^{(\mathrm{i})} \in \mathbb{R}$ such that

$$\sum_{\mathrm{P}\in\mathcal{F}} f^{(\mathrm{P},\mathrm{i})} + \lambda^{(\mathrm{i})} = f \ \text{ for all } \mathrm{i} \in \mathbb{N} \ \text{ and } \ \lim_{\mathrm{i}\to\infty} \lambda^{(\mathrm{i})} = \mathrm{f}^{\mathrm{DRLX}}.$$

Hence,

$$\lim_{\mathrm{i}\to\infty} \sum_{\mathrm{P}\in\mathcal{F}} f^{(\mathrm{P},\mathrm{i})} = f - \lim_{\mathrm{i}\to\infty} \lambda^{(\mathrm{i})} = f - \mathrm{f}^{\mathrm{DRLX}}.$$

$\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_{\mathrm{P}}$ is closed according to Corollary 2.7. Thus, $f - \mathrm{f}^{\mathrm{DRLX}} \in \sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_{\mathrm{P}}$. Assume now that there exists $\tilde{\mathrm{P}} \in \mathcal{F}$ for which the sequence $\{f^{(\tilde{\mathrm{P}},\mathrm{i})}\}_{\mathrm{i}\in\mathbb{N}}$ is unbounded, i.e. there exists a subsequence $\{f^{(\tilde{\mathrm{P}},\mathrm{i_j})}\}_{\mathrm{j}\in\mathbb{N}}$ such that $\lim_{\mathrm{j}\to\infty} ||f^{(\tilde{\mathrm{P}},\mathrm{i_j})}||_1 = \infty$. Then

$$\begin{aligned} p^{(\mathrm{j})} &:= \frac{f^{(\tilde{\mathrm{P}},\mathrm{i_j})}}{||f^{(\tilde{\mathrm{P}},\mathrm{i_j})}||_1} && \in \mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}}, \\ q^{(\mathrm{j})} &:= \frac{\sum_{\mathrm{P}\in\mathcal{F}\setminus\{\tilde{\mathrm{P}}\}} f^{(\mathrm{P},\mathrm{i_j})}}{||f^{(\tilde{\mathrm{P}},\mathrm{i_j})}||_1} && \in \sum_{\mathrm{P}\in\mathcal{F}\setminus\{\tilde{\mathrm{P}}\}} \mathcal{P}(\mathrm{K})_{\mathrm{P}} \end{aligned}$$

for all $\mathrm{j} \in \mathbb{N}$ and

$$\begin{aligned} \lim_{\mathrm{j}\to\infty} ||p^{(\mathrm{j})} + q^{(\mathrm{j})}||_1 &= \lim_{\mathrm{j}\to\infty} \frac{||f^{(\tilde{\mathrm{P}},\mathrm{i_j})} + \sum_{\mathrm{P}\in\mathcal{F}\setminus\{\tilde{\mathrm{P}}\}} f^{(\mathrm{P},\mathrm{i_j})}||_1}{||f^{(\tilde{\mathrm{P}},\mathrm{i_j})}||_1} \\ &= \lim_{\mathrm{j}\to\infty} \frac{||f - \lambda^{(\mathrm{i_j})}||_1}{||f^{(\tilde{\mathrm{P}},\mathrm{i_j})}||_1} \\ &= 0. \end{aligned}$$

The last equality follows from the fact that $\{\lambda^{(\mathrm{i_j})}\}_{\mathrm{j}\in\mathbb{N}}$ is bounded since $\lim_{\mathrm{j}\to\infty} \lambda^{(\mathrm{i_j})} = \mathrm{f}^{\mathrm{DRLX}}$. Thus, without loss of generality we can, after again extracting subsequences $\{\mathrm{j_k}\}_{\mathrm{k}\in\mathbb{N}}$, assume that

$$\begin{aligned} \lim_{\mathrm{k}\to\infty} p^{(\mathrm{j_k})} &= \quad p^\star \in \mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}}, \\ \lim_{\mathrm{k}\to\infty} q^{(\mathrm{j_k})} &= -p^\star \in \sum_{\mathrm{P}\in\mathcal{F}\setminus\{\tilde{\mathrm{P}}\}} \mathcal{P}(\mathrm{K})_{\mathrm{P}}. \end{aligned}$$

Hence,

$$p^\star \in \mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}} \cap \Big(- \sum_{\mathrm{P}\in\mathcal{F}\setminus\{\tilde{\mathrm{P}}\}} \mathcal{P}(\mathrm{K})_{\mathrm{P}} \Big) \subseteq \mathcal{P}(\mathrm{K}) \cap -\mathcal{P}(\mathrm{K}) \overset{(*)}{=} \{0\},$$

where $(*)$ follows form the assumption that K contains a full-dimensional ball. This is a contradiction to $||p^\star||_1 = \lim_{\mathrm{k}\to\infty} \frac{||f^{(\tilde{\mathrm{P}},\mathrm{j_k})}||_1}{||f^{(\tilde{\mathrm{P}},\mathrm{j_k})}||_1} = 1$ as the latter implies $p^\star \neq 0$. Thus, $\{f^{(\tilde{\mathrm{P}},\mathrm{i})}\}_{\mathrm{i}\in\mathbb{N}}$ is bounded. Consequently, we have shown that $\{(f^{(\mathrm{P},\mathrm{i})})_{\mathrm{P}\in\mathcal{F}}, \lambda^{(\mathrm{i})}\}_{\mathrm{i}\in\mathbb{N}}$ is also bounded. After extracting appropriate subsequences we can assume that

$$\lim_{\mathrm{i}\to\infty} f^{(\mathrm{P},\mathrm{i})} := f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}} \ \text{ for all } \ \mathrm{P} \in \mathcal{F}.$$

with $\sum_{\mathrm{P}\in\mathcal{F}} f^{\mathrm{P}} + \mathrm{f}^{\mathrm{DRLX}} = f$. □

2.3 Separation Problem

Different algorithmic approaches can be applied to solve (P-RLX'). One way is to use a cutting-plane algorithm that iteratively generates cuts for the sets $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$. We generate valid inequalities for $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$ from the maximization problem:

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

If $\mathbf{v}$ is not in $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$, then (SP) has a positive optimal value and the optimal solution $\mathbf{c}^*, \delta^*$ of (SP) yields an inequality $\langle \mathbf{c}^*, \mathbf{u} \rangle \geq \delta^*$, which is valid for all $\mathbf{u} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}}$ and violated for $\mathbf{u} = \mathbf{v}$. As a direct consequence of standard facts from convex optimization [16, Ch. 8.1.3] we obtain the following.

Proposition 2.12 ([9]).
Let P *be a pattern and* $\mathbf{v} \in \mathbb{R}^{\mathrm{P}}$. *Then the optimal value of* (SP) *is*

$$\operatorname{dist}\big(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \mathbf{v}\big) := \min_{\mathbf{u}\in\mathcal{M}(\mathrm{K})_{\mathrm{P}}} \|\mathbf{v} - \mathbf{u}\|_1.$$

△

The equality $\langle \mathbf{c}^*, \mathbf{u}\rangle = \delta^*$ defines a supporting hyperplane of $\mathcal{M}(\mathrm{K})_\mathrm{P}$ that contains a point of $\mathcal{M}(\mathrm{K})_\mathrm{P}$ closest to $\mathbf{v}$ in the ℓ_1-norm.

2.4 Shifting Patterns

The next two propositions are useful to generate new patterns by shifting existing ones.

Proposition 2.13 ([9]).
Let $\mathrm{P} \subseteq \mathbb{N}^\mathrm{n}$ *be a pattern with* $\operatorname{supp}(\mathrm{P}) \neq [\mathrm{n}]$ *and* $\xi \in \mathbb{N}^\mathrm{n}$ *a vector with* $\operatorname{supp}(\xi) \subseteq [\mathrm{n}] \setminus \operatorname{supp}(\mathrm{P})$*. Furthermore, let* $[\![\mathrm{K}]\!]_{\operatorname{supp}(\xi)}$ *be compact. Then*[5]

$$\mathcal{M}(\mathrm{K})_{\xi+\mathrm{P}} = \operatorname{cl}\operatorname{conv}\left(\underline{\mathbf{x}}^\xi_\mathrm{K}\, \mathcal{M}(\mathrm{K})_\mathrm{P} \cup \overline{\mathbf{x}}^\xi_\mathrm{K}\, \mathcal{M}(\mathrm{K})_\mathrm{P}\right). \qquad \triangle$$

Proof. Observe that $\underline{\mathbf{x}}^\xi_\mathrm{K}$ and $\overline{\mathbf{x}}^\xi_\mathrm{K}$ exist, since $[\![\mathrm{K}]\!]_{\operatorname{supp}(\xi)}$ is compact. The assertion follows from

$$\begin{aligned}
\mathcal{M}(\mathrm{K})_{\xi+\mathrm{P}} &= \operatorname{cl}\operatorname{conv}\left(\left\{(\mathbf{x}^\alpha)_{\alpha\in\xi+\mathrm{P}} : \mathbf{x} \in \mathrm{K}\right\}\right)\\
&= \operatorname{cl}\operatorname{conv}\left(\left\{(\mathbf{x}^{\xi+\beta})_{\beta\in\mathrm{P}} : \mathbf{x} \in \mathrm{K}\right\}\right)\\
&= \operatorname{cl}\operatorname{conv}\left(\left\{\mathbf{x}^\xi(\mathbf{x}^{\beta})_{\beta\in\mathrm{P}} : \mathbf{x} \in \mathrm{K}\right\}\right)
\end{aligned}$$

and the observation that $\mathbf{x}^\xi$ and $\mathbf{x}^\beta$ have no common factor, since $\operatorname{supp}(\xi)\cap\operatorname{supp}(\beta) = \emptyset$ for each $\beta \in \mathrm{P}$. Hence,

$$\begin{aligned}
\operatorname{conv}\left(\left\{\mathbf{x}^\xi(\mathbf{x}^{\beta})_{\beta\in\mathrm{P}} : \mathbf{x} \in \mathrm{K}\right\}\right) &= \operatorname{cl}\operatorname{conv}\left(\bigcup_{\mathbf{y}\in\mathrm{K}} \mathbf{y}^\xi\left\{(\mathbf{x}^{\beta})_{\beta\in\mathrm{P}} : \mathbf{x} \in \mathrm{K}\right\}\right)\\
&= \operatorname{cl}\operatorname{conv}\left(\left\{\mathrm{s}\, \mathcal{M}(\mathrm{K})_\mathrm{P} : \underline{\mathbf{x}}^\xi_\mathrm{K} \le \mathrm{s} \le \overline{\mathbf{x}}^\xi_\mathrm{K}\right\}\right)\\
&= \operatorname{cl}\operatorname{conv}\left(\underline{\mathbf{x}}^\xi_\mathrm{K}\, \mathcal{M}(\mathrm{K})_\mathrm{P} \cup \overline{\mathbf{x}}^\xi_\mathrm{K}\, \mathcal{M}(\mathrm{K})_\mathrm{P}\right). \qquad \square
\end{aligned}$$

[5] $\mathrm{s}\mathcal{M}(\mathrm{K})_\mathrm{P} = \{\mathrm{s}\mathbf{v} : \mathbf{v} \in \mathcal{M}(\mathrm{K})_\mathrm{P}\}$.

The dual of Proposition 2.13 is:

Proposition 2.14.

Let $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ *be a pattern with* $\operatorname{supp}(\mathrm{P}) \neq [\mathrm{n}]$ *and* $\xi \in \mathbb{N}^{\mathrm{n}}$ *a vector with* $\operatorname{supp}(\xi) \subseteq [\mathrm{n}] \setminus \operatorname{supp}(\mathrm{P})$. *Furthermore, let* $[\![\mathrm{K}]\!]_{\operatorname{supp}(\xi)}$ *be compact. Then*

$$\mathcal{P}(\mathrm{K})_{\xi+\mathrm{P}} = \Big\{ x^{\xi} g : \underline{\mathbf{x}}^{\xi}_{\mathrm{K}}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{P}} \text{ and } \overline{\mathbf{x}}^{\xi}_{\mathrm{K}}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{P}} \Big\}. \qquad \triangle$$

Proof. The assertion follows from

$$\begin{aligned}
\mathcal{P}(\mathrm{K})_{\xi+\mathrm{P}} &= \Big\{ f \in \mathcal{P}(\mathrm{K}) : f = \sum_{\beta \in \mathrm{P}} \mathrm{f}_{\xi+\beta} x^{\xi+\beta} \Big\} \\
&= \Big\{ f \in \mathcal{P}(\mathrm{K}) : f = x^{\xi} \sum_{\beta \in \mathrm{P}} \mathrm{f}_{\xi+\beta} x^{\beta} \Big\} \\
&= \Big\{ x^{\xi} g \in \mathcal{P}(\mathrm{K}) : g \in \mathbb{R}[x]_{\mathrm{P}} \Big\}.
\end{aligned}$$

Since $\operatorname{supp}(\xi) \cap \operatorname{supp}(\beta) = \emptyset$, $x^{\xi} g$ is nonnegative on K if and only if $\overline{\mathbf{x}}^{\xi}_{\mathrm{K}}\, g$ and $\underline{\mathbf{x}}^{\xi}_{\mathrm{K}}\, g$ are nonnegative on K. □

The shifting procedure has the following consequence for (SP):

Corollary 2.15 ([9]).

Let $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ *be a pattern with* $\operatorname{supp}(\mathrm{P}) \neq [\mathrm{n}]$ *and* $\xi \in \mathbb{N}^{\mathrm{n}}$ *a vector with* $\operatorname{supp}(\xi) \subseteq [\mathrm{n}] \setminus \operatorname{supp}(\mathrm{P})$. *Furthermore, let* $[\![\mathrm{K}]\!]_{\operatorname{supp}(\xi)}$ *be compact. The separation problem* (SP) *for the moment body* $\mathcal{M}(\mathrm{K})_{\xi+\mathrm{P}}$ *and a point* $\mathbf{v} \in \mathbb{R}^{\xi+\mathrm{P}}$ *can be formulated as follows*

$$\begin{array}{rlr}
\text{maximize} & \delta - \langle \mathbf{c}, \mathbf{v} \rangle & \\
\text{for} & \mathbf{c} \in \mathbb{R}^{\xi+\mathrm{P}} \text{ and } \delta \in \mathbb{R} & \\
\text{subject to} & \mathrm{c}_{\beta} \in [-1,1] & \text{for all } \beta \in \mathrm{P} \\
& \underline{\mathbf{x}}^{\xi}_{\mathrm{K}} \displaystyle\sum_{\beta \in \mathrm{P}} \mathrm{c}_{\xi+\beta} \mathbf{x}^{\beta} \geq \delta & \text{for all } \mathbf{x} \in \mathrm{K}, \\
& \overline{\mathbf{x}}^{\xi}_{\mathrm{K}} \displaystyle\sum_{\beta \in \mathrm{P}} \mathrm{c}_{\xi+\beta} \mathbf{x}^{\beta} \geq \delta & \text{for all } \mathbf{x} \in \mathrm{K}.
\end{array} \qquad \triangle$$

Proof. The assertion follows from applying Proposition 2.14 to (SP). □

Chapter 3

Known Convexification Techniques are Monomial Patterns

In this chapter we formulate established convexification techniques from the literature as patterns. These patterns have in common that their corresponding moment cones $\mathcal{C}(\mathrm{K})_{\mathrm{P}}$ or polynomial cones $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ admit an exact or approximate description that is computationally tractable. All pattern types introduced in this chapter can be used – alone or in combination – to generate computationally tractable pattern relaxations of (POP). The list of techniques is by no means exhaustive. Notably, in the context of moment and SOS relaxations there exists a wide range of related approaches that exploit intrinsic sparsity and symmetry, see [30, 53, 68, 67]. These techniques are not in conflict to the pattern relaxation approach and should also be considered in an implementation of the pattern relaxation.

Let $\alpha, \beta \in \mathbb{N}^{\mathrm{n}}$ be two vectors. We define

$$\mathbb{N}^{\mathrm{n}}_{\beta} := \{\gamma \in \mathbb{N}^{\mathrm{n}} : \gamma \leq \beta\}$$
$$\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}} := \{\gamma \in \mathbb{N}^{\mathrm{n}} : \gamma_1 + \cdots + \gamma_{\mathrm{n}} \leq 2\mathrm{d}\}$$

and denote the Hadamard product of α and β by $\alpha \circ \beta := (\alpha_{\mathrm{i}}\beta_{\mathrm{i}})_{\mathrm{i}\in[\mathrm{n}]}$. Whether we have a tractable description of the cones $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ and $\mathcal{C}(\mathrm{K})_{\mathrm{P}}$ depends not only on the pattern, but also on the set K. If we were to relax (POP) exclusively by patterns, for which we have a tractable description of $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$, the portfolio of pattern types at our disposal might be too small to find a pattern family that exploits the monomial structure of A to our satisfaction. As we are interested in lower bounds on the optimal value of (POP), there exist workarounds for this. The next remark shows how to deal with such situations.

Remark 3.1.
Let $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ *and* $\tilde{\mathrm{K}} \subseteq \mathbb{R}^{\mathrm{n}}$ *be closed semialgebraic sets such that* $\mathrm{K} \subseteq \tilde{\mathrm{K}}$ *and* K *is defined by the polynomials* $g^1, \ldots, g^{\mathrm{r}} \in \mathbb{R}[x]$*:*

$$\mathrm{K} = \left\{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^{\mathrm{i}}(\mathbf{x}) \geq 0 \ \text{ for all } \ \mathrm{i} \in [\mathrm{r}]\right\}. \tag{3.1}$$

A computationally tractable description of $\mathcal{P}(\tilde{\mathrm{K}})_{\mathrm{P}}$ *can be used to obtain a computationally tractable convex relaxation of* $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$*. This can be done in different ways, e.g.*

(C1) $\mathcal{P}(\tilde{\mathrm{K}})_{\mathrm{P}}$,

(C2) $\mathcal{P}(\tilde{\mathrm{K}})_{\mathrm{P}} + \sum_{\mathrm{i}\in[\mathrm{r}]} g^i \mathcal{P}(\tilde{\mathrm{K}})_{\mathrm{P}}$ *or*

(C3) $\sum_{\gamma\in\mathbb{N}^{\mathrm{r}}_{\beta}} \prod_{\mathrm{i}\in[\mathrm{r}]} (g^{\mathrm{i}})^{\gamma_{\mathrm{i}}} \mathcal{P}(\tilde{\mathrm{K}})_{\mathrm{P}}$,

where $\beta \in \mathbb{N}^{\mathrm{n}}$*. The cones* (C1) *to* (C3) *contain* $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$*. Hence, using* (C1)*,* (C2) *or* (C3) *to replace the respective cone* $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ *in* (D-RLX) *yields a valid relaxation of* (POP)*. While the idea of* (C1) *might seem trivial – clearly a polynomial that is nonnegative on* $\tilde{\mathrm{K}}$ *is also nonnegative on* K *– the global solver* ***BARON*** *considerably relies on it, see [50].* ***BARON*** *encases compact sets* K *by boxes* $\tilde{\mathrm{K}}$*, this allows the solver to rigorously use McCormick envelopes, for more details see Section 3.2 and Remark 3.6.* (C2) *and* (C3) *follow the principle 'sums and products of nonnegative polynomials are also nonnegative'. Note that* (C2) *is similar in spirit to quadratic modules [40, Ch. 2] and* (C3) *to Handelman's Positivstellensatz [40, Thm. 7.1.6] if* g^{i} *are linear. Both* (C2) *and* (C3)*, effectively 'create' new patterns, that are*

$$\tilde{\mathrm{P}} = \mathrm{P} \cup \bigcup_{\mathrm{i}\in[\mathrm{r}]} \big(\mathrm{supp}(g^{\mathrm{i}}) + \mathrm{P}\big) \text{ and } \tilde{\mathrm{P}} = \bigcup_{\gamma\in\mathbb{N}^{\mathrm{r}}_{\beta}} \Big(\sum_{\mathrm{i}\in[\mathrm{r}]} \mathrm{supp}((g^{\mathrm{i}})^{\gamma_{\mathrm{i}}}) + \mathrm{P}\Big), \text{ respectively.}$$

Furthermore, the cones (C1)*,*(C2) *and* (C3) *are, usually strictly, contained in* $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ *or* $\mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}}$ *Thus, using* (C1)*,*(C2) *or* (C3) *instead of* $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ *or* $\mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}}$ *within the pattern relaxation, one does not solve* (D-RLX)*, but a relaxation of* (D-RLX)*.* △

Plot Set Up 3.2.
For the illustration of the pattern types we consider different exponent sets for $\mathrm{n} = 2$ *in the following,*

$$\begin{aligned}
\mathrm{A}_1 &:= \{(0,0),(0,3),(2,0),(2,3)\},\\
\mathrm{A}_2 &:= \{(0,0),(1,1),(2,2),(3,3),(4,4),(5,5)\},\\
\mathrm{A}_3 &:= \{(0,4),(2,5),(2,8),(6,2)\},\\
\mathrm{A}_4 &:= \{(0,0),(2,2),(2,4),(4,2)\},
\end{aligned}$$

$$\begin{aligned}\mathrm{A}_5 &:= \{(0,2),(3,5),(6,8)\},\\ \mathrm{A}_6 &:= \{(2,8),(4,5),(6,2)\},\\ \mathrm{A}_{\mathrm{ex}} &:= \{(0,2),(1,1),(2,3),(2,4),(4,0),(5,5)\}.\end{aligned}$$

The exponent sets are used in different figures throughout the next chapters. The patterns and exponents are visualized as follows. The title of a subplot refers to the set of original exponents, which are depicted by red squares. The auxiliary exponents are depicted by blue dots. A pattern P *corresponds to an undirected smooth curve and all the colored points and squares that the curve passes through.*

Note that not all the features of certain pattern types can be visualized for $n = 2$*. For example in Section 3.2 it is not possible to visualize the difference between the pattern induced by McCormick envelopes and their generalization, i.e. multilinear patterns.*

The set A_4 *is the monomial support of the Motzkin polynomial:*

$$motz(x) = x_1^4x_2^2 + x_1^2x_2^4 - 3x_1^2x_2^2 + 1, \tag{3.2}$$

for which it is known that $motz(x) \geq 0$ *on* $\mathbb{R}^{\mathrm{n}}$ *and that* $motz(x) - \lambda$ *is for no* $\lambda \in \mathbb{R}$ *SOS, see [40, Prop. 1.2.2].*

Furthermore, A_{ex} *will also be used in the numerical result section.* △

3.1 Singletons

Definition 3.3 (Singleton Pattern).
We call $\{\alpha\}$ *with* $\alpha \in \mathbb{N}^{\mathrm{n}}$ *a singleton.* △

Singletons are the smallest patterns within the pattern framework. The moment body of a singleton $\{\alpha\}$ for a compact set K is the interval

$$\mathcal{M}(\mathrm{K})_{\{\alpha\}} = [\underline{\mathbf{x}}_{\mathrm{K}}^{\alpha}, \overline{\mathbf{x}}_{\mathrm{K}}^{\alpha}]$$

and the pattern relaxation of $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ induced by the family of singletons $\{\{\alpha\} : \alpha \in \mathrm{A}\}$ is $\mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}})$. While this seems rather unimpressive, the provided bounds on the monomial variables v_α can be exploited by branch-and-bound solvers [55] and in Algorithm 5.1 singletons ensure boundedness of (MP).

3.2 Multilinear Patterns

State-of-the-art solvers, like `BARON`, that are capable of solving (POP) globally for compact sets K, exploit multilinear structures in order to build their convex relaxations. The next definition formalizes these multilinear structures in terms of patterns.

Definition 3.4 (Multilinear Pattern).
Let $\alpha \in \mathbb{N}^n$ *and* $I \subseteq \{0,1\}^n$ *with* $I \neq \emptyset$*. We call*

$$\mathrm{ML}(\alpha, I) := \{(\alpha_1\omega_1, \dots, \alpha_n\omega_n) \in \mathbb{N}^n : \omega \in I\}$$

a multilinear pattern (ML). △

From this point onwards, whenever we introduce an operator for a pattern type, such as $\mathrm{ML}(\alpha, I)$, we imply that the input parameters satisfy the assumptions of the respective definition. We note that, depending on the context, the parameters may be named differently or compound expressions.

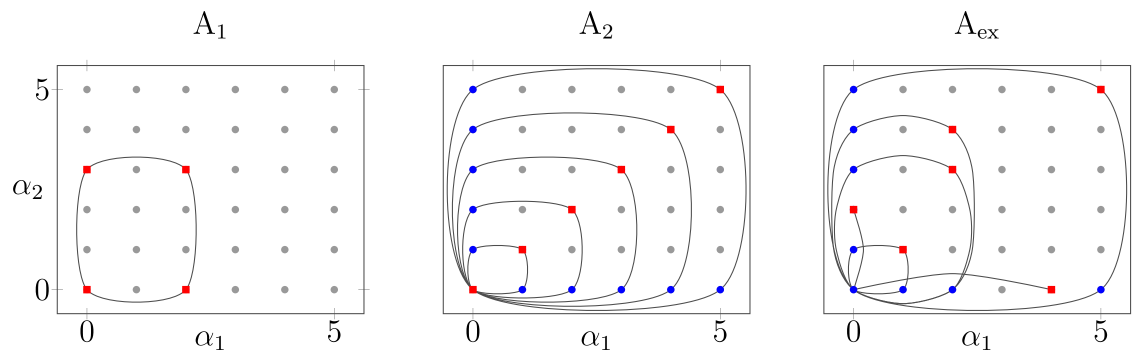

Figure 3.1: Visualization of possible multilinear patterns and families as described in Plot Set Up 3.2. **Left:** the pattern $\mathrm{ML}((2,3), \{0,1\}^2)$. **Middle:** $\{\mathrm{ML}(\alpha, \{0,1\}^2) : \alpha \in A_2\backslash\{\mathbf{0}\}\}$. **Right:** $\{\mathrm{ML}(\alpha, \{0,1\}^2) : \alpha \in A_{ex}\}$.

The multilinear structures that can be found in the literature usually correspond to rather static choices of I in $\mathrm{ML}(\alpha, I)$. Definition 3.4 poses a generalization of the multilinear structures that one usually encounters, in which the parameter I can be flexibly chosen. We briefly discuss this at the end of this section.

The notion 'multilinear' is coined by the observation that for $f \in \mathbb{R}[x]_{\mathrm{ML}(\alpha,I)}$ is multilinear in $y_i = x_i^{\alpha_i}$, i.e. $\sum_{\omega \in I} f_{\alpha\circ\omega} y^\omega \in \mathbb{R}[y]_I$ is a multilinear polynomial. It is a well-known fact that the convex envelope of a multilinear function over $K = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ is a polytope. In our context this implies the following:

Proposition 3.5 ([9]).
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l} < \mathbf{u}$ *and* $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$*. The moment body* $\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})}$ *is a polytope satisfying*

$$\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} = \mathrm{conv}(\mathrm{m}(\mathrm{V})_{\mathrm{I}})$$

with $\mathrm{V} := \{\underline{\mathbf{x}}_{\mathrm{K}}^{\alpha_1 \mathbf{e}^1}, \overline{\mathbf{x}}_{\mathrm{K}}^{\alpha_1 \mathbf{e}^1}\} \times \cdots \times \{\underline{\mathbf{x}}_{\mathrm{K}}^{\alpha_n \mathbf{e}^n}, \overline{\mathbf{x}}_{\mathrm{K}}^{\alpha_n \mathbf{e}^n}\}$. △

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

$$\begin{aligned} \mathrm{m}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} &= \{\mathrm{m}(\mathbf{x})_{\mathrm{ML}(\alpha,\mathrm{I})} : \mathbf{x} \in \mathrm{K}\} \\ &\cong \{\mathrm{m}((\mathbf{x}^{\alpha_i \mathbf{e}^i})_{\mathrm{i}\in[\mathrm{n}]})_{\mathrm{I}} : \mathbf{x} \in \mathrm{K}\} \\ &= \{\mathrm{m}(\mathbf{x})_{\mathrm{I}} : \mathbf{x} \in \tilde{\mathrm{K}}\} \\ &= \mathrm{m}(\tilde{\mathrm{K}})_{\mathrm{I}}. \end{aligned}$$

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

$$\mathrm{m}(\mathbf{x}^*)_{\mathrm{I}} = \frac{1}{2}\mathrm{m}(\mathbf{x}^\star - \varepsilon \mathbf{e}^{\mathrm{i}})_{\mathrm{I}} + \frac{1}{2}\mathrm{m}(\mathbf{x}^\star + \varepsilon \mathbf{e}^{\mathrm{i}})_{\mathrm{I}}.$$

This shows that the points of $\mathrm{m}(\tilde{\mathrm{K}} \setminus \mathrm{V})_{\mathrm{I}}$ are not extreme points of $\mathrm{m}(\tilde{\mathrm{K}})_{\mathrm{I}}$. Consequently, $\mathrm{m}(\tilde{\mathrm{K}})_{\mathrm{I}} = \mathrm{m}(\mathrm{V})_{\mathrm{I}}$ and therefore $\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} \cong \mathrm{conv}(\mathrm{m}(\mathrm{V})_{\mathrm{I}})$. Here '$\cong$' is used to emphasise that $\mathrm{m}(\tilde{\mathrm{K}})_{\mathrm{I}} \subseteq \mathbb{R}^{\mathrm{I}}$ and $\mathrm{m}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} \subseteq \mathbb{R}^{\mathrm{ML}(\alpha,\mathrm{I})}$ live in differently indexed spaces. □

A illustration of Proposition 3.5 is given in Figure 3.2. For the rest of this section we assume that $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ for $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ with $\mathbf{l} < \mathbf{u}$.

Figure 3.1 shows three different applications of multilinear patterns. The left subfigure illustrates how one pattern connects all exponents of an exponent set, here A_1, directly. The corresponding moment body is shown in Figure 3.2 for $\mathrm{K} = [0, 1]^2$. The middle subplot is an example of an exponent set A_2, where a sparse application of multilinear patterns, as suggested in [26, 11], does not connect any of the original exponents[1]. While `BARON` usually handles polynomial problems with sparse monomial support well, the numerical results in Chapter 6 show that it struggles with problems that have a monomial support that are chain shaped like A_2. The right subplot is an example where a sparse application of multilinear patterns, that is the family

$$\{\mathrm{ML}(\alpha, \{0, 1\}^2) : \alpha \in \mathrm{A}\},$$

[1]Recall that the **0** exponent does not connect monomial variables, see Definition 2.4.

indirectly connects the original exponents $(2,3)$ and $(2,4)$ via the auxiliary variable $(2,0)$.

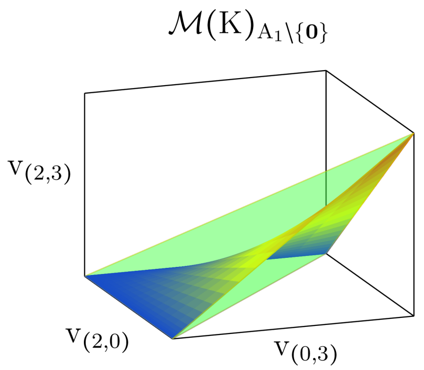

Figure 3.2: The moment body $\mathcal{M}(\mathrm{K})_{\mathrm{A}_1\setminus\{\mathbf{0}\}}$ in green and the surface $\{\mathrm{m}(\mathbf{x})_{\mathrm{A}_1\setminus\{\mathbf{0}\}} : \mathbf{x} \in \mathrm{K}\}$ for $\mathrm{K} = [0,1]^2$. Note that $\mathcal{M}([0,1]^2)_{\mathrm{A}_1} = \{\mathbf{1}\} \times \mathcal{M}([0,1]^2)_{\mathrm{A}_1\setminus\{\mathbf{0}\}}$.

Remark 3.6.

1. *From Proposition 3.5 it immediately follows that* $\mathcal{P}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})}$ *for* $\mathrm{K} = \mathrm{Box}(\mathbf{l},\mathbf{u})$ *is an polyhedral cone: Lemma 2.6 asserts that a polynomial* $f \in \mathbb{R}[x]_{\mathrm{ML}(\alpha,\mathrm{I})}$ *is nonnegative on* K *if and only if* $\langle \mathbf{f}, \mathbf{v}\rangle \geq 0$ *for all* $\mathbf{v} \in \mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})}$*. Thus,*
$$\mathcal{P}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} = \Big\{ f \in \mathbb{R}[x]_{\mathrm{ML}(\alpha,\mathrm{I})} : \sum_{\omega\in\mathrm{I}} \mathrm{f}_{\alpha\circ\omega}\tilde{\mathrm{x}}^{\omega} \geq 0 \ \text{ for all } \ \tilde{\mathbf{x}} \in \mathrm{V} \Big\}.$$

2. *If* K *is not an axis-parallel box, but just a compact set, the description given in Proposition 3.5 may still be used to derive a relaxation of* $\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})}$*. Observe that* $\mathrm{K} \subseteq \mathrm{Box}\left(\underline{\mathbf{x}}_{\mathrm{K}}^{\{\mathrm{e}^{\mathrm{i}}:\mathrm{i}\in[\mathrm{n}]\}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\{\mathrm{e}^{\mathrm{i}}:\mathrm{i}\in[\mathrm{n}]\}}\right)$*. Thus,*
$$\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})} \subseteq \mathcal{M}\Big(\mathrm{Box}\left(\underline{\mathbf{x}}_{\mathrm{K}}^{\{\mathrm{e}^{\mathrm{i}}:\mathrm{i}\in[\mathrm{n}]\}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\{\mathrm{e}^{\mathrm{i}}:\mathrm{i}\in[\mathrm{n}]\}}\right)\Big)_{\mathrm{ML}(\alpha,\mathrm{I})}.$$
△

Corollary 3.7 ([9]).
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l} < \mathbf{u}$ *and* $\mathrm{K} = \mathrm{Box}(\mathbf{l},\mathbf{u})$*. The separation problem for the moment body* $\mathcal{M}(\mathrm{K})_{\mathrm{ML}(\alpha,\mathrm{I})}$ *and a point* $\mathbf{v} \in \mathbb{R}^{\mathrm{ML}(\alpha,\mathrm{I})}$ *can be formulated as the following linear*

program

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

where $\mathrm{V} := \{\underline{\mathbf{x}}_\mathrm{K}^{\alpha_1 \mathbf{e}^1}, \overline{\mathbf{x}}_\mathrm{K}^{\alpha_1 \mathbf{e}^1}\} \times \cdots \times \{\underline{\mathbf{x}}_\mathrm{K}^{\alpha_n \mathbf{e}^n}, \overline{\mathbf{x}}_\mathrm{K}^{\alpha_n \mathbf{e}^n}\}$. △

Proof. The assertion follows from Proposition 3.5.

The linear problem in Corollary 3.7 involves at most $2^{\#\,\mathrm{supp}(\alpha)}$ inequalities indexed by vectors $\mathbf{w} \in \mathrm{m}(\mathrm{V})_\mathrm{I}$.

Algorithm 3.1: Recursive McCormick Envelopes

Input: $\alpha \in \mathbb{N}^n$ with $\#\,\mathrm{supp}(\alpha) \geq 2$
Output: family of multilinear patterns $\mathcal{F}_{\mathrm{rec}}^\alpha$

(0) Set $\mathrm{J} = \{\alpha\}$ and $\mathcal{F}_{\mathrm{rec}}^\alpha = \emptyset$.
(1) While $\mathrm{J} \neq \emptyset$:

Choose $\beta \in \mathrm{J}$.

Decompose β into $\beta', \beta'' \in \mathbb{N}^n$ such that $\beta = \beta' + \beta''$, $\mathrm{supp}(\beta') \neq \emptyset, \mathrm{supp}(\beta'') \neq \emptyset$ and $\mathrm{supp}(\beta') \cap \mathrm{supp}(\beta'') = \emptyset$.

Remove β from J and add the multilinear pattern $\{\beta, \beta', \beta'', \mathbf{0}\}$ to $\mathcal{F}_{\mathrm{rec}}^\alpha$.

If $\#\,\mathrm{supp}(\beta') \geq 2$, add β' to J.

If $\#\,\mathrm{supp}(\beta'') \geq 2$, add β'' to J.

Multilinear patterns for multilinear polynomials

Multilinear patterns can be found in different contexts in the literature. In their basic version they are used to convexify multilinear polynomials. An essential building block for the convexification of product terms is the McCormick envelope [42], that is the convexification of bilinear products $\mathrm{x}_1\mathrm{x}_2$ by a tight description of the moment body $\mathcal{M}(\mathrm{K})_{\mathrm{ML}((1,1),\{0,1\}^2)}$. McCormick envelopes have been successfully used to build convex relaxations of multilinear monomials by applying them recursively. Algorithm 3.1

describes this recursion for a monomial $\mathbf{x}^\alpha$. This procedure corresponds to a binary tree with root α and the moment body of each pattern in $\mathcal{F}^\alpha_{\mathrm{rec}}$ is tightly described by a McCormick envelope. In general it is not clear how to favorably decompose an exponent $\beta \in \mathrm{J}$ in Algorithm 3.1. For the smallest nontrivial case $\#\operatorname{supp}(\beta) = 3$ this has been investigated in [63].

Another way to convexify $\mathbf{x}^\alpha$ with $\alpha \in \{0,1\}^{\mathrm{n}}$ and $\#\operatorname{supp}(\alpha) \geq 2$ is to introduce for each factor $\mathrm{x}_\mathrm{i}^{\alpha_\mathrm{i}}$ with $\alpha_\mathrm{i} \neq 0$ a moment variable $\mathrm{v}_{\alpha_\mathrm{i}}$ [20]. This corresponds to the pattern

$$\mathrm{P}^\alpha := \mathrm{ML}(\alpha, \{\mathbf{1}, \mathbf{0}\} \cup \{\mathbf{e}^\mathrm{i} : \mathrm{i} \in \operatorname{supp}(\alpha)\}). \tag{3.3}$$

Multilinear Patterns for general Exponent Sets

Multilinear patterns have also been applied to polynomials $f \in \mathbb{R}[x]_\mathrm{A}$ with general exponent set $\mathrm{A} \subseteq \mathbb{N}^\mathrm{n}$ and compact K. [11, 26]. We outline the underlying concept of this: Let

$$\begin{aligned}
&\Gamma := \left\{\alpha_i \mathbf{e}^\mathrm{i} : \alpha \in \mathrm{A} \ \text{ and } \ \mathrm{i} \in [\mathrm{n}]\right\} \backslash \left\{\mathbf{0}\right\} \quad \text{and} \\
&\tilde{\mathrm{A}} := \Big\{\beta \in \{0,1\}^\Gamma : \text{ there exists } \ \alpha \in \mathrm{A} \ \text{ with } \ \textstyle\sum_{\gamma\in\Gamma} \beta_\gamma \gamma = \alpha \\
&\qquad\qquad \text{and } \ \beta_\gamma + \beta_\omega \leq 1 \ \text{ for all } \ \gamma, \omega \in \Gamma \ \text{ with } \ \operatorname{supp}(\gamma) \cap \operatorname{supp}(\omega) + \emptyset\Big\}.
\end{aligned}$$

Then using the substitution $y_{\alpha_\mathrm{i}\mathbf{e}^\mathrm{i}} = x^{\alpha_i \mathbf{e}^i}$ a multilinear intermediate[2]

$$\tilde{f} = \sum_{\beta\in\tilde{\mathrm{A}}} \mathrm{f}_{\Gamma\beta} y^\beta \in \mathbb{R}[y]_{\tilde{\mathrm{A}}}$$

of f is generated. The multilinear intermediate $\tilde{f}$ has exponents from $\tilde{\mathrm{A}}$ and the indeterminates y are indexed by Γ.

Example 3.8.
For $\mathrm{A}_{\mathrm{ex}} = \{(0,2), (1,1), (2,3), (2,4), (4,0), (5,5)\}$ *the sets* Γ *and* $\tilde{\mathrm{A}}$ *from above are*

$$\Gamma = \{(0,1), (0,2), (0,3), (0,4), (0,5), (1,0), (2,0), (4,0), (5,0)\}$$

and

$$\begin{aligned}
\tilde{\mathrm{A}} = \{&(0,1,0,0,0,0,0,0,0), (1,0,0,0,0,1,0,0,0), (0,0,1,0,0,0,1,0,0), \\
&(0,0,0,1,0,0,1,0,0), (0,0,0,0,0,0,0,1,0), (0,0,0,0,1,0,0,0,1)\}.
\end{aligned}$$
△

[2]Here we use $\Gamma\beta := \sum_{\gamma\in\Gamma} \gamma\beta_\gamma$ similar to the matrix-vector product.

This corresponds to relaxing the usually non-polyhedral $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ with the polytope $\mathcal{M}(\mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}))_{\tilde{\mathrm{A}}}$ and (POP) by

$$\begin{array}{rl} \text{minimize} & \langle \tilde{\mathbf{f}}, \mathbf{v} \rangle \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\tilde{\mathrm{A}}} \\ \text{subject to} & \mathbf{v} \in \mathcal{M}\big(\mathrm{Box}\left(\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}\right)\big)_{\tilde{\mathrm{A}}}. \end{array} \tag{3.4}$$

The problem (3.4) is then further relaxed using Algorithm 3.1 or (3.3). This entire process can also be expressed using multilinear patterns, see for example subfigure A_{ex} in Figure 3.1.

For $\mathrm{A} = \{\alpha\}$ with $\alpha \in \mathbb{N}^{\mathrm{n}}$ the pattern relaxation corresponding to the pattern family $\{\mathrm{P}^{\alpha}\}$ is tight, while relaxation corresponding to $\mathcal{F}_{\mathrm{rec}}^{\alpha}$ is usually not tight for $\mathrm{A} = \{\alpha\}$. It is however not clear which system

$$\mathbf{v}_{\mathrm{P}^{\alpha}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}^{\alpha}} \quad \text{for all} \;\; \alpha \in \mathrm{A} \;\; \text{with} \;\; \#\,\mathrm{supp}(\alpha) \geq 2 \tag{3.5}$$

or

$$\mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}} \quad \text{for all} \;\; \mathrm{P} \in \mathcal{F}_{\mathrm{rec}}^{\alpha}, \;\; \alpha \in \mathrm{A} \;\; \text{with} \;\; \#\,\mathrm{supp}(\alpha) \geq 2 \tag{3.6}$$

yields a tighter convex relaxation of $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ for $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ with $\#\mathrm{A} \geq 2$. This is due to the different choice of auxiliary variables and how the original moment variables are connected by the different pattern families in (3.5) and (3.6). In our Definition 3.4 of $\mathrm{ML}(\alpha, \mathrm{I})$ the parameter I allows to flexibly choose auxiliary variables and thereby control the connective properties of the of multilinear pattern family. In doing so one can combine the merits of P^{α} and $\mathcal{F}_{\mathrm{rec}}^{\alpha}$.

3.3 Expression Trees

Convexification using expression trees is common in general nonlinear optimization [59, 62] over compact sets K. In this section K is an axis-parallel box.[3] This approach is based on the observation that each algebraic expression is made up of a certain set of elementary operations, such as powers, linear combinations, or products of expressions. A decomposition of an algebraic expression into these operations can be visualized using an algebraic expression tree, like in Figure 3.3. This is a rooted tree with nodes labeled by terms occurring in the expression. Each term is built up from

[3] This is not much of a restriction, as for this technic K is usually approximated by a box, see [50].

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. For example, the term

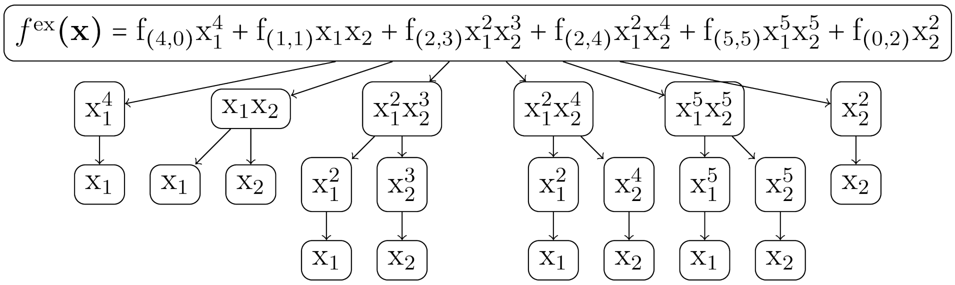

Figure 3.3: A possible algebraic expression tree for the polynomial f^{ex} with $\mathrm{supp}(f^{\mathrm{ex}}) \subseteq \mathrm{A}_{\mathrm{ex}}$ and the set A_{ex} from Plot Set Up 3.2.

$\mathrm{x}_1^2\mathrm{x}_2^3$ in Figure 3.3 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 $\mathrm{v}_{(2,3)}$, $\mathrm{v}_{(2,0)}$ and $\mathrm{v}_{(0,3)}$, respectively. The relationship of these variables is captured by the pattern $\mathrm{P} = \{(2,3),(2,0),(0,3)\}$ and the corresponding moment body $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$ is described by the well-known McCormick inequalities. The variable $\mathrm{v}_{(0,3)}$ is further connected to $\mathrm{v}_{(0,1)}$ by exponentiation. The corresponding pattern is $\{(0,1),(0,3)\}$. All patterns induced by the tree in Figure 3.3 are visualized in Figure 3.4.

Remark 3.9.
There exist other ways to form an expression tree. In Figure 3.3 the product $\mathrm{x}_1^2\mathrm{x}_2^3$ *could also be split into* $\mathrm{x}_1^2\mathrm{x}_2^2$ *and* x_2 *or* x_1 *and* $\mathrm{x}_1\mathrm{x}_2^3$*. However, the corresponding patterns* $\{(2,3),(2,2),(0,1)\}$ *and* $\{(2,3),(1,0),(1,3)\}$ *and their moment bodies are not tightly described McCormick envelopes anymore if* K *is an axis-parallel box.* △

Since expression trees normally correspond to patterns of small size, they lead to weak, but efficiently computable relaxations, which are often used in divide-and-conquer approaches like branch-and-bound methods. 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.

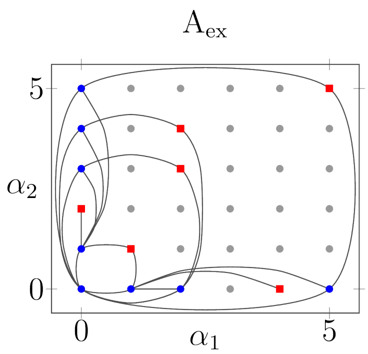

Figure 3.4: Visualization of the pattern family $\mathcal{F}$ induced by the expression tree from Figure 3.3 as described in Plot Set Up 3.2. The induced family is $\mathcal{F} := \{\mathrm{ML}(\alpha, \{0,1\}^2) : \alpha = (1,1),(2,3),(2,4),(5,5)\} \cup \{\{(0,1),(0,\mathrm{i})\} : \mathrm{i} \in [5]\backslash\{1\}\} \cup \{\{(1,0),(\mathrm{i},0)\} : \mathrm{i} \in \{2,4,5\}\}$.

3.4 Bound-Factor Products

Another convexification approach is based on so-called bound-factor products [19]. This approach is also tailored for $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ with $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ and $\mathbf{l} < \mathbf{u}$ [4]. Since the polynomials $x_{\mathrm{i}} - \mathrm{l}_{\mathrm{i}}$ and $\mathrm{u}_{\mathrm{i}} - x_{\mathrm{i}}$ are nonnegative on K, the products of these polynomials (with repetitions allowed) are also nonnegative on K. So, one can consider the products

$$F^{\alpha,\beta}(x) := \prod_{\mathrm{i}\in[\mathrm{n}]} (x_{\mathrm{i}} - \mathrm{l}_{\mathrm{i}})^{\alpha_{\mathrm{i}}-\beta_{\mathrm{i}}}(\mathrm{u}_{\mathrm{i}} - x_{\mathrm{i}})^{\beta_{\mathrm{i}}} \tag{3.7}$$

of $|\alpha|$ polynomials with α_{i} linear factors depending on the variable x_{i}, where $\alpha, \beta \in \mathbb{N}^{\mathrm{n}}$ and $\alpha \geq \beta$. For a generic choice of $\mathbf{l}$ and $\mathbf{u}$, the polynomial $F^{\alpha,\beta}(\mathbf{x})$ includes all monomials with exponents in the pattern

$$\mathrm{BF}(\alpha) := \{0, \dots, \alpha_1\} \times \cdots \times \{0, \dots, \alpha_{\mathrm{n}}\}.$$

By substituting $v_\gamma = x^\gamma$ for all $\gamma \in \mathrm{BF}(\alpha)$ we obtain a linearization $LF^{\alpha,\beta}(\mathbf{v})$ of $F^{\alpha,\beta}(x)$. The system of linear inequalities

$$LF^{\alpha,\beta}(v) \geq 0 \ \text{ for all } \ \beta \in \mathrm{BF}(\alpha) \tag{3.8}$$

is valid for $\mathbf{v} \in \mathcal{M}(\mathrm{K})_{\mathrm{BF}(\alpha)}$. This approach can also be viewed as hierarchical since one can increase the order of the bound-factor products in order to tighten the relaxation.

[4] Though using Remark 3.6 this approach can be applied to situations, where K is compact but not a box.

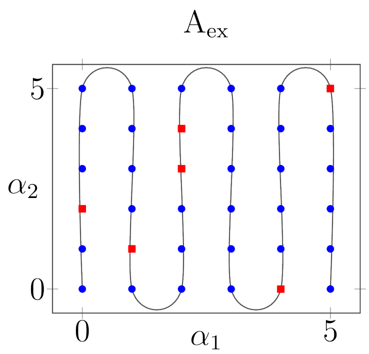

Figure 3.5: Visualization of the bound-factor product pattern $\mathrm{BF}((5,5))$ applied to A_{ex} as described in Plot Set Up 3.2.

In the polynomial optimization community the method of bound-factor products is known as the dual of Handelman's hierarchy [27]. Within this approach one groups monomial variables into patterns of a rather large 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_{ex} from Plot Set Up 3.2 one is forced to use at least one pattern $\mathrm{BF}(\alpha)$ with $\alpha_1 \geq 5$ and $\alpha_2 \geq 5$, which means that at least 36 monomial variables have to be introduced, see Figure 3.5. Another issue is that the system of linear inequalities (3.8) is not a tight description of $\mathcal{M}(\mathrm{K})_{\mathrm{BF}(\alpha)}$. To mitigate this, these kinds of relaxations have also been used within branch-and-bound strategies.

3.5 Moment Relaxations

The most popular convexification techniques in the polynomial optimization community are the moment relaxation and its dual counterpart, the SOS relaxation [2, 38, 40]. This approach introduces a large number of monomial variables and links them all with one large pattern using PSD 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 in practice good approximation properties at the expense of large semidefinite programs (SDPs), see [49] for computational studies. Even the lowest possible hierarchy level of the moment relaxation for medium-sized problems results in huge SDPs. For example a generic 6-variate polynomial of degree 12 the SDP already includes a dense PSD constraint of size 924. However, strategies exist to make the approach more tractable, e.g., [1, 69, 36] and [2, Ch. 8].

To derive a so-called moment relaxation of (POP), the following representation of the moment body $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ in terms of probability measures is used [38, Sec. 4.2]:

$$\mathcal{M}(\mathrm{K})_{\mathrm{A}} = \left\{ \int \mathrm{m}(\mathbf{x})_{\mathrm{A}} \mu(d\mathbf{x}) : \mu \text{ is a probability measure on K} \right\}.$$

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

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

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

Definition 3.10 (Moment Matrix and Localizing Matrix [37, Ch.2.7.1]). *The localizing matrix* $\mathbf{M}_{\mathrm{k}}(g, \mathbf{v})$ *for a polynomial* g *with coefficients* $(\mathrm{g}_\alpha)_\alpha$ *and the moment matrix* $\mathbf{M}_{\mathrm{k}}(\mathbf{v})$ *are defined as*

$$\mathbf{M}_{\mathrm{k}}(g, \mathbf{v}) := \left(\sum_{\gamma \in \mathbb{N}^{\mathrm{n}}} \mathrm{g}_\gamma \mathrm{v}_{\gamma+\alpha+\beta} \right)_{\alpha,\beta \in \mathbb{N}^{\mathrm{n}}_{\mathrm{k}}}$$

$$\mathbf{M}_{\mathrm{k}}(\mathbf{v}) := \mathbf{M}_{\mathrm{k}}(1, \mathbf{v}).$$

△

Theorem 3.11 ([37, Thm. 2.44]).
Let $g^1, \ldots, g^{\mathrm{r}} \in \mathbb{R}[x]$ *be* n*-variate polynomials such that there exist SOS polynomials* $s^1, \ldots, s^{\mathrm{r}}$ *for which*

$$\left\{ \mathbf{x} \in \mathbb{R}^{\mathrm{n}} : s^0(\mathbf{x}) + \sum_{\mathrm{i} \in [\mathrm{r}]} s^{\mathrm{i}}(\mathbf{x}) g^{\mathrm{i}}(\mathbf{x}) \geq 0 \right\} \tag{3.9}$$

is compact. Furthermore, let

$$\mathrm{K} = \left\{ \mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^{\mathrm{i}}(\mathbf{x}) \geq 0 \ \text{ for all } \ \mathrm{i} \in [\mathrm{r}] \right\}. \tag{3.10}$$

A sequence $\{\mathrm{v}_\alpha\}_{\alpha \in \mathbb{N}^{\mathrm{n}}}$ *has a finite Borel representing measure with support in* K *if and* *only if*

$$\begin{aligned} \mathbf{M}_{\mathrm{k}}(\mathbf{v}) &\in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{k}}} \ \text{ and} \\ \mathbf{M}_{\mathrm{k}}(g^{\mathrm{i}}, \mathbf{v}) &\in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{k}}} \ \text{ for all } \ \mathrm{i} \in [\mathrm{r}] \ \text{ and for all } \ \mathrm{k}. \end{aligned} \qquad \triangle$$

Remark 3.12.
If $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ *with* $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *and* $\mathbf{l} < \mathbf{u}$*, the box* K *can be described by the polynomials* $g^{\mathrm{i}}(x) := (x_{\mathrm{i}} - \mathrm{l}_{\mathrm{i}})(\mathrm{u}_{\mathrm{i}} - x_{\mathrm{i}})$ *for* $\mathrm{i} \in [\mathrm{n}]$*, i.e.* $\mathrm{K} = \{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^{\mathrm{i}}(\mathbf{x}) \geq 0 \ \text{ for all } \ \mathrm{i} \in [\mathrm{n}]\}$*. Note that the assumptions of Theorem 3.11 hold for these* $g^1, \dots, g^{\mathrm{n}}$*.* $\triangle$

With Theorem 3.11 we can formulate (C-POP) as

$$\begin{array}{rlll} \text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle & & \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathbb{N}^{\mathrm{n}}} & & \\ \text{subject to} & \mathbf{M}_{\mathrm{k}}(\mathbf{v}) & \in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{k}}} & \text{for all } \ \mathrm{k}, \\ & \mathbf{M}_{\mathrm{k}-2}(g^{\mathrm{i}}, \mathbf{v}) & \in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{k}-2}} & \text{for all } \ \mathrm{i} \in [\mathrm{r}] \ \text{ and for all } \ \mathrm{k}, \\ & \mathrm{v}_{\mathbf{0}} = 1. & & \end{array}$$

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

$$\begin{array}{rlll} \text{minimize} & \langle \mathbf{f}, \mathbf{v} \rangle & & \\ \text{for} & \mathbf{v} \in \mathbb{R}^{\mathbb{N}^{\mathrm{n}}_{2}} & & \\ \text{subject to} & \mathbf{M}_{\mathrm{d}}(\mathbf{v}) & \in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}}, & \\ & \mathbf{M}_{\mathrm{d}-2}(g^{\mathrm{i}}, \mathbf{v}) & \in \mathbb{S}_+^{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}-2}} & \text{for all } \ \mathrm{i} \in [\mathrm{r}], \\ & \mathrm{v}_{\mathbf{0}} = 1 & & \end{array}$$

is a lower bound on the optimal value of (POP). This problem has one SDP constraint of size $\binom{\mathrm{n}+\mathrm{d}}{\mathrm{d}}$ that involves the variables v_α with $\alpha \in \mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}$ and n SDP constraints of size $\binom{\mathrm{n}+\mathrm{d}-1}{\mathrm{d}-1}$ that involve the moment variables v_α with $\alpha \in \mathbb{N}^{\mathrm{n}}_{2\mathrm{d}-2}$. Hence, the moment relaxation uses the pattern $\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}$. Note that for general problems it is not possible to reduce the size of the mentioned SDP constraints [7].

For a small example like A_{ex} from Plot Set Up 3.2 with $\deg(\mathrm{A}_{\mathrm{ex}}) = 10$ and $\mathrm{n} = 2$

this adds up to 66 moment variables. Figure 3.6 shows the pattern corresponding to the lowest hierarchy level that is needed in order to solve (POP) for $\mathrm{A} = \mathrm{A}_{\mathrm{ex}}$ and $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ with $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ and $\mathbf{l} < \mathbf{u}$. This hierarchy level involves an SDP constraint with a 21×21 matrix.

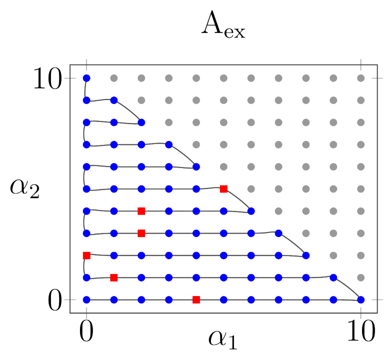

Figure 3.6: Visualization of pattern $\mathbb{N}^2_{10}$ corresponding to the lowest hierarchy level of the moment relaxation applied to A_{ex} as described in Plot Set Up 3.2.

3.6 Nonnegative Circuit Polynomials

A relatively new convexification technique utilizes so-called sums of nonnegative circuit polynomials (SONCs) [29] to build convex relaxations of (POP). In their standard form these relaxations are employed for unconstrained problems. We therefore assume in this section that $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$. However, using Remark 3.1 SONCs can also be adapted to derive lower bounds for constrained optimization problems. SONCs have been successfully used to find the minimum of polynomials that are not SOS like the Motzkin polynomial (3.2). The current literature regarding SONC includes: relative entropy relaxations of SONC cones [17], experimental evaluations using geometric programming (GP) [58], characterizations of the dual of the SONC cone [31] and linear approximations [22].

The smallest building block of SONC is the following pattern type.

Definition 3.13 (Circuit).
Let $\mathrm{S} \subseteq 2\mathbb{N}^{\mathrm{n}}$ be an affinely independent set of cardinality k with $2 \leq \mathrm{k} \leq \mathrm{n}+1$ and let $\mu^{\beta} \in \mathbb{R}^{\mathrm{S}}_{++}$ be a weight vector with $\sum_{\alpha \in \mathrm{S}} \mu^{\beta}_{\alpha} = 1$ such that the corresponding convex combination of exponents in S satisfies

$$\beta := \sum_{\alpha \in \mathrm{S}} \mu^{\beta}_{\alpha} \alpha \in \mathbb{N}^{\mathrm{n}}.$$

We call

$$\mathrm{CR}(\mathrm{S}, \mu^\beta) := \mathrm{S} \cup \{\beta\}$$

a circuit (CR), β the center of the circuit and the elements of $\mathbb{R}[x]_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ circuit polynomials.[5] *Furthermore, let $\mathcal{F}$ be a family of circuits. Modifying the notation established in [29], we refer to a polynomial f that satisfies*

$$f \in \sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{P}} \tag{3.11}$$

as a sum of nonnegative circuit polynomials (SONC) and the Minkowski sum of cones $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ as a SONC cone with respect to $\mathcal{F}$. △

This slightly extends the original definition of the SONC cone by the family $\mathcal{F}$. Furthermore, in this definition the SONC cone is closed, whereas this is not the case for the one given in [29, Def. 1.3], that is

$$\Big\{ \sum_{\mathrm{P}\in\mathcal{F}} f^{\mathrm{P}} \in \mathbb{R}[x]_{\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}} : \mathcal{F} \text{ is a family of circuits and } f^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \\ \text{satisfies } \mathrm{f}_\beta^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \neq 0,\ \mathbf{f}_{\mathrm{S}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} > \mathbf{0} \text{ for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F} \Big\}.$$

Definition 3.14 (Power Cone).
Let $\mathrm{S} \subseteq \mathbb{N}^{\mathrm{n}}$ be a nonempty and finite set with cardinality $\mathrm{k} := \#\mathrm{S}$, $\beta \in \mathbb{N}^{\mathrm{n}}$ with $\beta \notin \mathrm{S}$ and $\mu \in \mathbb{R}^{\mathrm{S}}_{++}$ with $\sum_{\alpha\in\mathrm{S}} \mu_\alpha = 1$. The $\mathrm{k}+1$ dimensional power cone $\mathbb{P}_\mu$ corresponding to μ is defined as

$$\mathbb{P}_\mu := \Big\{ (\mathbf{u}_{\mathrm{S}}, \mathrm{u}_\beta) \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : |\mathrm{u}_\beta| \leq \prod_{\alpha\in\mathrm{S}} \mathrm{u}_\alpha^{\mu_\alpha} \Big\}.$$

△

In the following we derive the known characterization of $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ from its dual $\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$. We will see that $\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ and $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ have a native representation by the power cone and its dual. These cones can be exploited computationally by conic programming solvers. Since the power cone representation does not involve a nonlinear transformation, like the one given in [29], it allows to freely combine circuits with all other pattern types. We also give an alternative proof for [29, Thm. 5.5].

[5] As always, whenever we write $\mathrm{CR}(\mathrm{S}, \mu^\beta)$ we imply that the parameters S and μ^β satisfy the assumptions of Definition 3.13.

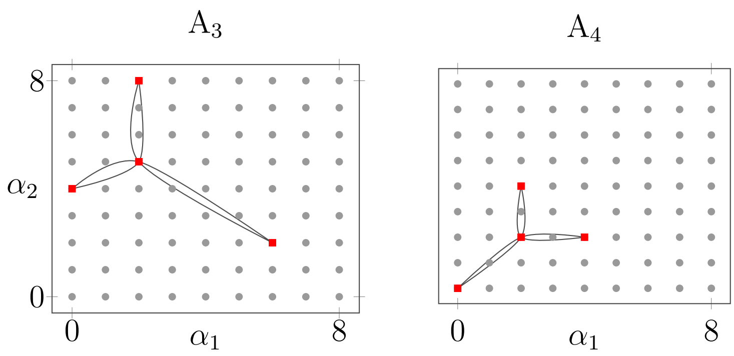

Figure 3.7: Visualization of two circuits as described in Plot Set Up 3.2. **Left:** $\mathrm{CR}(\mathrm{S}, \mu)$ with $\mathrm{S} = \{(0,4), (2,8), (6,2)\}$ and $\mu = (0.6, 0.3, 0.1)$. **Right:** monomial support of the Motzkin polynomial $motz(x) = x_1^4 x_2^2 + x_1^2 x_2^4 - 3x_1^2 x_2^2 + 1$.

Theorem 3.15 (Dual Power Cone, [18, Thm. 4.3.1.]).
Let $\mathrm{S} \subseteq \mathbb{N}^{\mathrm{n}}$ *a nonempty and finite set,* $\beta \in \mathbb{N}^{\mathrm{n}}$ *with* $\beta \notin \mathrm{S}$, $\mu \in \mathbb{R}^{\mathrm{S}}_{++}$ *with* $\sum_{\alpha \in \mathrm{S}} \mu_\alpha = 1$ *and*[6]

$$\mathbf{B}^\mu := \begin{pmatrix} \mathrm{diag}(\mu) & 0 \\ 0 & 1 \end{pmatrix} \in \mathbb{R}^{(\mathrm{S} \cup \{\beta\}) \times (\mathrm{S} \cup \{\beta\})}.$$

The dual of the power cone $\mathbb{P}_\mu$ *is*

$$\mathbb{P}^*_\mu = \mathbf{B}^\mu \cdot \mathbb{P}_\mu$$

or equivalently

$$\mathbb{P}^*_\mu = \left\{ (\mathbf{w}_\mathrm{S}, \mathrm{w}_\beta) \in \mathbb{R}^{\mathrm{S}}_+ \times \mathbb{R}^{\{\beta\}} : |\mathrm{w}_\beta| \leq \prod_{\alpha \in [\mathrm{S}]} \left(\frac{\mathrm{w}_\alpha}{\mu_\alpha} \right)^{\mu_\alpha} \right\}.$$

△

For more information on the power cone the interested reader is referred to [18] and the `MOSEK` Modeling Cookbook [4]. In [18] the power cone is thoroughly investigated and self-concordant barrier functions for the power cone are presented. These barrier functions make the power cone accessible to conic programming solvers that employ interior-point methods. The `MOSEK` Modeling Cookbook provides tips for the usage of power cones and documents the current state-of-the-art algorithms used for solving power cone formulations.

[6] $\mathrm{diag}(\mu)$ is a $\mathrm{S} \times \mathrm{S}$ diagonal matrix with diagonal $\mu \in \mathbb{R}^{\mathrm{S}}_{++}$.

Theorem 3.16.
Let $\mathrm{CR}(\mathrm{S},\mu^\beta)$ *be a circuit. If* $\beta \in 2\mathbb{N}^{\mathrm{n}}$ *then*

$$\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : 0 \le \mathrm{v}_\beta \le \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\Big\}$$

and if $\beta \notin 2\mathbb{N}^{\mathrm{n}}$ *then*

$$\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : |\mathrm{v}_\beta| \le \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\Big\}. \qquad \triangle$$

Proof. Let

$$\begin{aligned}\mathcal{K}^{\ge} &:= \operatorname{cl}\operatorname{cone}\Big(\Big\{\ \mathrm{m}(\mathbf{x})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathbb{R}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} : \mathbf{x} \in \mathbb{R}^{\mathrm{n}} \text{ and } \mathbf{x}^\beta \ge 0\Big\}\Big)\\ \mathcal{K}^{\le} &:= \operatorname{cl}\operatorname{cone}\Big(\Big\{\ \mathrm{m}(\mathbf{x})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathbb{R}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} : \mathbf{x} \in \mathbb{R}^{\mathrm{n}} \text{ and } \mathbf{x}^\beta \le 0\Big\}\Big).\end{aligned}$$

We claim that for $\beta \in \mathbb{N}^{\mathrm{n}}$

$$\mathcal{K}^{\ge} = \Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : 0 \le \mathrm{v}_\beta \le \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\Big\} \tag{3.12}$$

and that for $\beta \notin 2\mathbb{N}^{\mathrm{n}}$

$$\mathcal{K}^{\le} = \Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : 0 \ge \mathbf{v}_\beta \ge -\prod_{\alpha\in\mathrm{S}} \mathbf{v}_\alpha^{\mu_\alpha^\beta}\Big\}. \tag{3.13}$$

After proving the claim, the assertion of the theorem follows for $\beta \in 2\mathbb{N}^{\mathrm{n}}$ since $\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \mathcal{K}^{\ge}$ and for $\beta \notin 2\mathbb{N}^{\mathrm{n}}$ since $\mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \mathcal{K}^{\le} \cup \mathcal{K}^{\ge}$. We start by showing (3.12).

Since $\mathrm{S} \subseteq 2\mathbb{N}^{\mathrm{n}}$, $\mathrm{m}(\mathbf{x})_{\mathrm{S}} \ge \mathbf{0}$ holds for all $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ and therefore $\mathbf{v}_{\mathrm{S}} \ge \mathbf{0}$ for all $\mathbf{v} \in \mathcal{C}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$. Hence, writing $\mathbf{x}^\beta$ as $\prod_{\alpha\in\mathrm{S}}(\mathrm{x}^\alpha)^{\mu_\alpha^\beta}$ yields

$$\begin{aligned}\mathcal{K}^{\ge} &= \operatorname{cl}\operatorname{cone}\Big(\Big\{\ \mathrm{m}(\mathbf{x})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathbb{R}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} : \mathbf{x} \in \mathbb{R}^{\mathrm{n}} \text{ and } \prod_{\alpha\in\mathrm{S}}(\mathrm{x}^\alpha)^{\mu_\alpha^\beta} \ge 0\Big\}\Big)\\ &= \operatorname{cl}\operatorname{cone}\Big(\Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : \mathbf{v}_{\mathrm{S}} = \mathrm{m}(\mathbf{x})_{\mathrm{S}}, \mathbf{x} \in \mathbb{R}^{\mathrm{n}} \text{ and } \mathrm{v}_\beta = \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\Big\}\Big) \qquad (3.14)\\ &= \operatorname{cl}\operatorname{cone}\Big(\Big\{\mathbf{v} \in \mathbb{R}^{\mathrm{S}}_{+} \times \mathbb{R}^{\{\beta\}} : \mathrm{v}_\beta = \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\Big\}\Big). \qquad (3.15)\end{aligned}$$

To see the latter equality, observe that (3.14) is contained in (3.15). For the other inclusion it is enough to show that for all $\mathbf{v}_{\mathrm{S}} \in \mathbb{R}_+^{\mathrm{S}}$ it holds

$$\left(\mathbf{v}_{\mathrm{S}}, \prod_{\alpha \in \mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\right) \in \mathcal{K}^{\geq}. \tag{3.16}$$

To this end, for each $\omega \in \mathrm{S}$ we provide sequences $\left\{\lambda^{(\mathrm{k},\omega)}\right\}_{\mathrm{k}\in\mathbb{N}\setminus\{0\}} \subseteq \mathbb{R}_+$ and $\left\{\mathbf{x}^{(\mathrm{k},\omega)}\right\}_{\mathrm{k}\in\mathbb{N}\setminus\{0\}} \subseteq \mathbb{R}^{\mathrm{n}}$ such that the product of $\lambda^{(\mathrm{k},\omega)}$ and $\mathrm{m}\left(\mathbf{x}^{(\mathrm{k},\omega)}\right)_{\mathrm{S}}$ converges to the unit vector $\mathbf{e}^{\omega,\mathrm{S}} \in \mathbb{R}^{\mathrm{S}}$, i.e.

$$\lim_{\mathrm{k}\to\infty} \lambda^{(\mathrm{k},\omega)}\, \mathrm{m}\left(\mathbf{x}^{(\mathrm{k},\omega)}\right)_{\mathrm{S}} = \mathbf{e}^{\omega,\mathrm{S}}. \tag{3.17}$$

Since the elements in S are affinely independent, there exists for each $\omega \in \mathrm{S}$ a vector $\mathbf{p} \in \mathbb{R}^{\mathrm{n}}$ with $\mathbf{p} \neq \mathbf{0}$ and $\langle \mathbf{p}, \omega \rangle > \langle \mathbf{p}, \alpha \rangle$ for all $\alpha \in \mathrm{S}\setminus\{\omega\}$. Using $\mathbf{p}$ we define $\lambda^{(\mathrm{k},\omega)} := \mathrm{k}^{-\langle \mathbf{p},\omega\rangle}$ and $\mathbf{x}^{(\mathrm{k},\omega)} := (\mathrm{k}^{\mathrm{p_i}})_{\mathrm{i}\in[\mathrm{n}]}$. Then

$$\begin{aligned}
\lim_{\mathrm{k}\to\infty} \lambda^{(\mathrm{k},\omega)} \left(\mathbf{x}^{(\mathrm{k},\omega)}\right)^{\omega} &= \lim_{\mathrm{k}\to\infty} \mathrm{k}^{-\langle \mathbf{p},\omega\rangle} \prod_{\mathrm{i}\in[\mathrm{n}]} \mathrm{k}^{\mathrm{p_i}\omega_i} \\
&= \lim_{\mathrm{k}\to\infty} \mathrm{k}^{\langle \mathbf{p},\omega\rangle - \langle \mathbf{p},\omega\rangle} \\
&= 1
\end{aligned}$$

and for $\alpha \in \mathrm{S}\setminus\{\omega\}$

$$\begin{aligned}
\lim_{\mathrm{k}\to\infty} \lambda^{(\mathrm{k},\omega)} \left(\mathbf{x}^{(\mathrm{k},\omega)}\right)^{\alpha} &= \lim_{\mathrm{k}\to\infty} \mathrm{k}^{-\langle \mathbf{p},\omega\rangle} \prod_{\mathrm{i}\in[\mathrm{n}]} \mathrm{k}^{\mathrm{p_i}\alpha_i} \\
&= \lim_{\mathrm{k}\to\infty} \mathrm{k}^{\langle \mathbf{p},\alpha\rangle - \langle \mathbf{p},\omega\rangle} \\
&= 0,
\end{aligned}$$

since $\langle \mathbf{p}, \alpha \rangle - \langle \mathbf{p}, \omega \rangle < 0$. This shows (3.17). For each $\mathbf{v} \in \mathbb{R}_+^{\mathrm{S}}$

$$\mathbf{v}^{(\mathrm{k})} := \sum_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha \lambda^{(\mathrm{k},\alpha)}\, \mathrm{m}\left(\mathbf{x}^{(\mathrm{k},\alpha)}\right)_{\mathrm{S}}$$

converges for $\mathrm{k} \to \infty$ to $\mathbf{v}$ and

$$\lim_{\mathrm{k}\to\infty} \prod_{\alpha\in\mathrm{S}} \left(\mathrm{v}_\alpha^{(\mathrm{k})}\right)^{\mu_\alpha^\beta} = \prod_{\alpha\in\mathrm{S}} \left(\lim_{\mathrm{k}\to\infty} \mathrm{v}_\alpha^{(\mathrm{k})}\right)^{\mu_\alpha^\beta} = \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}.$$

Observe that from (3.14) follows that $\{(\mathbf{v}^{(\mathrm{k})}, \prod_{\alpha\in \mathrm{S}}(\mathrm{v}_\alpha^{(\mathrm{k})})^{\mu_\alpha^\beta})\}_{\mathrm{k}\in\mathbb{N}\setminus\{0\}}$ is a sequence in $\mathcal{K}^{\geq}$. Hence, we have shown (3.16) since $\mathcal{K}^{\geq}$ is a convex and closed cone and since this sequence converges to $(\mathbf{v}_\mathrm{S}, \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta})$.

Next, let $\mathcal{K}$ be the right hand side of (3.12). From the observation that $\mathbf{v}_\mathrm{S} \mapsto \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}$ is concave on $\mathbb{R}^\mathrm{S}_{\geq 0}$ follows that $\mathcal{K}$ is convex. Note that $\mathcal{K}$ is a closed cone. Hence, with (3.15) follows $\mathcal{K}^{\geq} \subseteq \mathcal{K}$. For the reverse inclusion observe that from (3.15) it follows $\mathbf{e}^{\alpha,\mathrm{S}\cup\{\beta\}} \in \mathcal{K}^{\geq}$ for all $\alpha \in \mathrm{S}$. This shows that $(\mathbf{v}_\mathrm{S}, 0) \in \mathcal{K}^{\geq}$ for every $\mathbf{v}_\mathrm{S} \in \mathbb{R}^\mathrm{S}_+$ since $\mathcal{K}^{\geq}$ is a convex cone. Since $(\mathbf{v}_\mathrm{S}, \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}) \in \mathrm{K}^{\geq}$ this implies due to the convexity of $\mathcal{K}^{\geq}$ that $(\mathbf{v}_\mathrm{S}, \mathrm{v}_\beta) \in \mathcal{K}^{\geq}$ for all $0 \leq \mathrm{v}_\beta \leq \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}$. This shows $\mathcal{K} \subseteq \mathcal{K}^{\geq}$ and therefore the claim (3.12).

Repeating the above steps, this time using the convexity of $\mathbf{v}_\mathrm{S} \mapsto -\prod_{\alpha\in\mathrm{S}} \mathbf{v}_\alpha^{\mu_\alpha^\beta}$ on $\mathbb{R}^\mathrm{S}_+$, we obtain the claim (3.13). $\square$

Theorem 3.16 shows that the native cones to represent $\mathcal{C}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ are power cones:

$$\begin{aligned} \mathcal{C}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} &= \mathbb{P}_{\mu^\beta} \cap \left(\mathbb{R}^\mathrm{S} \times \mathbb{R}^{\{\beta\}}_+\right) \quad && \text{for } \beta \in 2\mathbb{N}^\mathrm{n} \\ \mathcal{C}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} &= \mathbb{P}_{\mu^\beta} && \text{for } \beta \notin 2\mathbb{N}^\mathrm{n}. \end{aligned} \tag{3.18}$$

Furthermore, the proof of Theorem 3.16 can be modified to show that for $\beta \in \mathbb{N}^\mathrm{n}$

$$\mathcal{C}(\mathbb{R}^\mathrm{n}_+)_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \left\{\mathbf{v} \in \mathbb{R}^\mathrm{S}_+ \times \mathbb{R}^{\{\beta\}} : 0 \leq \mathrm{v}_\beta \leq \prod_{\alpha\in\mathrm{S}} \mathrm{v}_\alpha^{\mu_\alpha^\beta}\right\}. \tag{3.19}$$

For that simply note that $\mathcal{C}(\mathbb{R}^\mathrm{n}_+)_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \mathcal{K}^{\geq}$. With the above we can derive the known characterization of $\mathcal{P}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ from [29, Thm. 1.1] by simply dualizing $\mathcal{C}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$.

Corollary 3.17 ([29, Thm. 1.1]).
We define the circuit number Θ_f for a polynomial $f \in \mathbb{R}[x]_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ as

$$\Theta_f := \prod_{\alpha\in\mathrm{S}} \left(\frac{\mathrm{f}_\alpha}{\mu_\alpha^\beta}\right)^{\mu_\alpha^\beta}.$$

The cone of nonnegative circuit polynomials with support $\mathrm{CR}(\mathrm{S},\mu^\beta)$ *is for* $\beta \notin 2\mathbb{N}^\mathrm{n}$

$$\mathcal{P}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \left\{g \in \mathbb{R}[x] : g = \sum_{\alpha\in\mathrm{S}} \mathrm{g}_\alpha x^\alpha + \mathrm{g}_\beta x^\beta, \mathbf{g}_\mathrm{S} \geq \mathbf{0} \text{ and } |\mathrm{g}_\beta| \leq \Theta_g\right\}$$

and for $\beta \in 2\mathbb{N}^{\mathrm{n}}$

$$\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR(S},\mu^\beta)} = \Big\{ g \in \mathbb{R}[x] : g = \sum_{\alpha \in \mathrm{S}} \mathrm{g}_\alpha x^\alpha + \mathrm{g}_\beta x^\beta, \mathbf{g}_{\mathrm{S}} \geq \mathbf{0} \text{ and } \mathrm{g}_\beta \geq -\Theta_g \Big\}. \qquad \triangle$$

Proof. $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR(S},\mu^\beta)}$ is a closed and convex cone. Thus, the assertion follows from (3.18) and Theorem 3.15. That is for $\beta \notin 2\mathbb{N}^{\mathrm{n}}$

$$\begin{aligned} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR(S},\mu^\beta)} &= \mathcal{C}(\mathbb{R}^{\mathrm{n}})^*_{\mathrm{CR(S},\mu^\beta)} \\ &= \mathbb{P}^*_{\mu^\beta} \\ &= \Big\{ (\mathbf{g}_{\mathrm{S}}, \mathrm{g}_\beta) \in \mathbb{R}^{\mathrm{S}}_+ \times \mathbb{R}^{\{\beta\}} : |\mathrm{g}_\beta| \leq \Theta_g \Big\} \end{aligned}$$

and for $\beta \in 2\mathbb{N}^{\mathrm{n}}$

$$\begin{aligned} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CR(S},\mu^\beta)} &= \mathcal{C}(\mathbb{R}^{\mathrm{n}})^*_{\mathrm{CR(S},\mu^\beta)} \\ &= \Big(\mathbb{P}_{\mu^\beta} \cap \Big(\mathbb{R}^{\mathrm{S}} \times \mathbb{R}^{\{\beta\}}_+ \Big) \Big)^* \\ &= \mathbb{P}^*_{\mu^\beta} + \Big(\big\{ \mathbf{0}^{\mathrm{S}} \big\} \times \mathbb{R}^{\{\beta\}}_+ \Big) \\ &= \Big\{ (\mathbf{g}_{\mathrm{S}}, \mathrm{g}_\beta) \in \mathbb{R}^{\mathrm{S}}_+ \times \mathbb{R}^{\{\beta\}} : \mathrm{g}_\beta \in -\Theta_g + \mathbb{R}_+ \Big\}. \end{aligned} \qquad \square$$

In the following we give an alternative proof of [29, Thm. 5.5]. To this end we use the following conventions: for a family of circuits $\mathcal{F}$

- $\mathrm{Z}^{\mathrm{e}} := \{\beta \in 2\mathbb{N}^{\mathrm{n}} : \mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F}\}$ is the set of even centers of $\mathcal{F}$,
- $\mathrm{Z}^{\mathrm{u}} := \{\beta \notin 2\mathbb{N}^{\mathrm{n}} : \mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F}\}$ is the set of odd centers of $\mathcal{F}$,
- $\mathrm{Z} \ := \mathrm{Z}^{\mathrm{e}} \cup \mathrm{Z}^{\mathrm{u}}$ is the set of all centers of $\mathcal{F}$.

Corresponding to this we define

- $\mathcal{F}^{\mathrm{e}} := \{\mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F} : \beta \in \mathrm{Z}^{\mathrm{e}}\}$,
- $\mathcal{F}^{\mathrm{u}} := \{\mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F} : \beta \in \mathrm{Z}^{\mathrm{u}}\}$.

Hence, $\mathcal{F} = \mathcal{F}^{\mathrm{e}} \cup \mathcal{F}^{\mathrm{u}}$. As usual, we use $\mathrm{f}_\alpha^{\mathrm{CR(S},\mu^\beta)} = \mathrm{vec}(f^{\mathrm{CR(S},\mu^\beta)})_\alpha$ for $\alpha \in \mathrm{CR}(\mathrm{S}, \mu^\beta)$ and $\mathrm{f}_\alpha^{\mathrm{CR(S},\mu^\beta)} = 0$ for $\alpha \notin \mathrm{CR}(\mathrm{S}, \mu^\beta)$. Let $f \in \mathbb{R}[x]_{\mathrm{A}}$ be a polynomial, $\mathcal{F}$ be a family of circuits with $\mathrm{A} \subseteq \mathrm{A}_{\mathcal{F}}$ and $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$. With Corollary 3.17 we formulate the optimization program (D-RLX) as

$$
\begin{array}{lrll}
\text{maximize} & \lambda & & \\
\text{for} & \lambda & \in \mathbb{R}, & \\
& \mathbf{f}_{\mathrm{S}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} & \in \mathbb{R}_+^{\mathrm{S}} & \text{for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}, \\
& \mathrm{f}_{\beta}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} & \in \mathbb{R} & \text{for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F} \\
\text{subject to} & \prod_{\alpha\in\mathrm{S}} \left(\frac{\mathrm{f}_{\alpha}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}}{\mu_\alpha^\beta}\right)^{\mu_\alpha^\beta} & \geq |\mathrm{f}_{\beta}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}| & \text{for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}^{\mathrm{u}}, \\
& -\prod_{\alpha\in\mathrm{S}} \left(\frac{\mathrm{f}_{\alpha}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}}{\mu_\alpha^\beta}\right)^{\mu_\alpha^\beta} & \leq \mathrm{f}_{\beta}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} & \text{for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}^{\mathrm{e}}, \\
& \sum_{\mathrm{CR}(\mathrm{S},\mu^\beta)\in\mathcal{F}} \mathrm{f}_{\alpha}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} & = \mathrm{f}_\alpha & \text{for all } \alpha \in \mathrm{A}_{\mathcal{F}}\backslash\{\mathbf{0}\}, \\
& \sum_{\mathrm{CR}(\mathrm{S},\mu^\beta)\in\mathcal{F}} \mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} & = \mathrm{f}_{\mathbf{0}} - \lambda. &
\end{array}
\tag{D-SONC}
$$

Lemma 3.18.
Let $f \in \mathbb{R}[x]_{\mathrm{A}}$. If $\mathcal{F}$ is a family of circuits that satisfies

(A1) $\mathrm{A} = \mathrm{A}_{\mathcal{F}}$,

(A2) $\mathbf{f}_{\mathrm{S}\backslash\{\mathbf{0}\}} > \mathbf{0}$ *for all* $\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}$,

(A3) $\mathbf{0} \in \mathrm{S}$ *for all* $\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}$,

then the optimal value of (D-SONC) *is attained.* △

Proof. For the proof we provide a generic feasible solution $\left(\mathbf{f}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}\right)_{\mathrm{CR}(\mathrm{S},\mu^\beta)\in\mathcal{F}}$ and λ for (D-SONC). By choosing

$$
\mathrm{f}_{\alpha}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \frac{\mathrm{f}_\alpha}{\#\{\mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}}) \in \mathcal{F} : \alpha \in \mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}})\}}
$$

for all $\alpha \in \mathrm{CR}(\mathrm{S},\mu^\beta)\backslash\{\mathbf{0}\}$ and for all $\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}$ we ensure that the constraints $\sum_{\mathrm{CR}(\mathrm{S},\mu^\beta)\in\mathcal{F}} \mathrm{f}_{\alpha}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} = \mathrm{f}_\alpha$ for all $\alpha \in \mathrm{A}_{\mathcal{F}}\backslash\{\mathbf{0}\}$ are satisfied. From assumption (A2) follows for this choice that

$$
\mathbf{f}_{\mathrm{S}\backslash\{\mathbf{0}\}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} > \mathbf{0} \ \text{ for all } \mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}. \tag{3.20}
$$

Due to assumption (A3) and (3.20) we can make

$$\prod_{\alpha\in\mathrm{S}}\left(\frac{\mathrm{f}_\alpha^{\mathrm{CR}(\mathrm{S},\mu^\beta)}}{\mu_\alpha^\beta}\right)^{\mu_\alpha^\beta}$$

arbitrarily large by choosing $\mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ large enough. Hence, by choosing an appropriate $\mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ we can satisfy the inequality constraints of (D-SONC). For the last step, set

$$\lambda = \mathrm{f}_{\mathbf{0}} - \sum_{\mathrm{CR}(\mathrm{S},\mu^\beta)\in\mathcal{F}} \mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}. \tag{3.21}$$

This shows that (D-SONC) is feasible. Replacing the objective function with (3.21) shows that the objective value of (D-SONC) is decreasing in $\mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)}$. Due to the constraint $\mathbf{f}_{\mathrm{S}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathbb{R}_+^{\mathrm{S}}$ in (D-SONC), we have $\mathrm{f}_{\mathbf{0}}^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \geq 0$. This implies that the objective value of (D-SONC) is bounded by $\mathrm{f}_{\mathbf{0}}$. With that, Lemma 2.11 yields the assertion. □

Theorem 3.19 ([29, Thm. 5.5 + Cor. 7.4]).
Let $f \in \mathbb{R}[x]_{\mathrm{A}}$ and $\mathcal{F}$ satisfy the assumptions (A1), (A2), (A3) *and*

(A4) $\mathrm{f}_\beta \in \mathbb{R}\backslash\{0\}$ *for all* $\beta \in \mathrm{Z}^{\mathrm{u}}$,

(A5) $\mathrm{f}_\beta < 0$ *for all* $\beta \in \mathrm{Z}^{\mathrm{e}}$,

(A6) *for all* $\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}$ *holds* $\beta \notin \mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}})$ *for all* $\mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}}) \in \mathcal{F}$ *with* $\mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}}) \neq \mathrm{CR}(\mathrm{S},\mu^\beta)$,

(A7) $\mathrm{S} = \tilde{\mathrm{S}}$ *for all* $\mathrm{CR}(\mathrm{S},\mu^\beta), \mathrm{CR}(\tilde{\mathrm{S}},\mu^{\tilde{\beta}}) \in \mathcal{F}$,

(A8) *there exists* $\tilde{\mathbf{x}} \in \mathbb{R}^{\mathrm{n}}$ *such that* $\mathrm{f}_\beta\tilde{\mathbf{x}}^\beta < 0$ *for all* $\beta \in \mathrm{Z}$.

Furthermore, let $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$. *Then the optimal values* $\mathrm{f}^{\mathrm{pop}}$ *of* (POP) *and* $\mathrm{f}^{\mathrm{sonc}}$ *of* (D-SONC) *coincide.* △

The first 3 steps of our proof of Theorem 3.19 are similar to the steps of the proofs given in [29].

Proof. First we make some observations. Lemma 3.18 asserts that there exists $\lambda \in \mathbb{R}$ such that $f - \lambda \geq 0$ on $\mathbb{R}^{\mathrm{n}}$. Let $\lambda \in \mathbb{R}^{\mathrm{n}}$ be an arbitrary scalar such that $g := f - \lambda \geq 0$ on $\mathbb{R}^{\mathrm{n}}$. From (A7) and Definition 3.13 follows that $\mathrm{Z} \subseteq \mathrm{relint}(\mathrm{conv}(\mathrm{S}))$. Thus, g has the form

$$g = \sum_{\alpha \in \mathrm{S}} \mathrm{g}_\alpha x^\alpha + \sum_{\beta \in \mathrm{Z}} \mathrm{g}_\beta x^\beta, \tag{3.22}$$

where $\mathrm{g}_\alpha = \mathrm{f}_\alpha$ for all $\alpha \in \mathrm{A}\backslash\{\mathbf{0}\}$ and $\mathrm{g}_\mathbf{0} = \mathrm{f}_\mathbf{0} - \lambda$. Furthermore, since we always assume that $\mathrm{A} \neq \emptyset$, $\mathcal{F}$ must contain at least one circuit $\mathrm{CR}(\mathrm{S}, \mu^\beta)$. By Definition 3.13 holds that $\mathrm{S} \geq 2$ and therefore $\#\mathrm{Z} \geq 1$. Assumption (A6) states that each center $\beta \in \mathrm{Z}$ is contained in exactly one $\mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F}$. That means

$$\mathcal{F} = \{\mathrm{CR}(\mathrm{S}, \mu^\beta) : \beta \in \mathrm{Z}\}.$$

The rest of the proof consists of five steps:

1. Proving that we can assume that g has the form

$$g = \sum_{\alpha \in \mathrm{S}} \mathrm{g}_\alpha x^\alpha + \sum_{\beta \in \mathrm{Z}} \mathrm{g}_\beta x^\beta \text{ with } \mathbf{g}_\mathrm{S} > \mathbf{0} \text{ and } \mathbf{g}_\mathrm{Z} < 0. \tag{3.23}$$

2. Proving that $g \geq 0$ on $\mathbb{R}^\mathrm{n}$ if and only if $g \geq 0$ on $\mathbb{R}^\mathrm{n}_+$.

3. Showing that we can assume that S satisfies

$$\mathrm{S} \subseteq \{\mathrm{r}\mathbf{e}^\mathrm{i} : \mathrm{i} \in [\mathrm{n}] \text{ and } \mathrm{r} \in 2\mathbb{N}\} \cup \{\mathbf{0}\}, \tag{3.24}$$

   i.e. the exponents in S are located on the coordinate axes.

4. Proving that the optimal value $\mathrm{f}^{\mathrm{pop}}$ of (POP) is attained.

5. At last we use duality to finish the proof.

1.) Using $\tilde{\mathbf{x}}$ from assumption (A8), let $\mathrm{sign}(\tilde{\mathbf{x}}) := (\mathrm{sign}(\tilde{\mathrm{x}}_\mathrm{i}))_{\mathrm{i}\in[\mathrm{n}]}$ and let $\tilde{g}(x) := g((\mathrm{sign}(\tilde{\mathrm{x}}_\mathrm{i})x_\mathrm{i})_{\mathrm{i}\in[\mathrm{n}]})$. Note that for every $\mathbf{x} \in \mathbb{R}^\mathrm{n}$ there exists a unique vector $\bar{\mathbf{x}} = (\mathrm{sign}(\tilde{\mathrm{x}}_\mathrm{i})\mathrm{x}_\mathrm{i})_{\mathrm{i}\in[\mathrm{n}]} \in \mathbb{R}^\mathrm{n}$. Thus, $g \geq 0$ on $\mathbb{R}^\mathrm{n}$ if and only if $\tilde{g} \geq 0$ on $\mathbb{R}^\mathrm{n}$. Next, the assumptions (A2), (A4) and (A5) imply the following[7]

$$\begin{aligned}
\mathrm{vec}(\tilde{g})_\alpha &= \mathrm{g}_\alpha \,\mathrm{sign}(\tilde{\mathbf{x}})^\alpha = \quad \mathrm{g}_\alpha > 0 && \text{for all } \alpha \in \mathrm{S}\backslash\{\mathbf{0}\},\\
\mathrm{vec}(\tilde{g})_\beta &= \mathrm{g}_\beta \,\mathrm{sign}(\tilde{\mathbf{x}})^\beta = \quad \mathrm{g}_\beta < 0 && \text{for all } \beta \in \mathrm{Z}^{\mathrm{e}},\\
\mathrm{vec}(\tilde{g})_\beta &= \mathrm{g}_\beta \,\mathrm{sign}(\tilde{\mathbf{x}})^\beta = \quad \mathrm{g}_\beta < 0 && \text{for all } \beta \in \mathrm{Z}^{\mathrm{u}} \text{ with } \mathrm{g}_\beta < 0 \text{ and}\\
\mathrm{vec}(\tilde{g})_\beta &= \mathrm{g}_\beta \,\mathrm{sign}(\tilde{\mathbf{x}})^\beta = -\mathrm{g}_\beta < 0 && \text{for all } \beta \in \mathrm{Z}^{\mathrm{u}} \text{ with } \mathrm{g}_\beta > 0.
\end{aligned}$$

Observe that if $\mathrm{vec}(\tilde{g})_{\mathbf{0}} \leq 0$, then the Newton polytope of $\tilde{g}$ has at least one vertex[8] with a negative coefficient. This contradicts $\tilde{g} \geq 0$ on $\mathbb{R}^{\mathrm{n}}$. Thus, $\mathrm{vec}(\tilde{g})_{\mathrm{Z}} < \mathbf{0}$ and $\mathrm{vec}(\tilde{g})_{\mathrm{S}} > \mathbf{0}$. Hence, we assume from now on that g has the form (3.23).

2.) Let $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$. We define $|\mathbf{x}| := (|\mathrm{x}_1|, \dots, |\mathrm{x}_{\mathrm{n}}|) \in \mathbb{R}^{\mathrm{n}}_+$ for a vector $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$. Recall that $\mathrm{S} \subseteq 2\mathbb{N}^{\mathrm{n}}$. This implies $\mathrm{g}_\alpha|\mathbf{x}|^\alpha = \mathrm{g}_\alpha\mathbf{x}^\alpha$ for each $\alpha \in \mathrm{S}$. Furthermore, $\mathrm{g}_\beta|\mathbf{x}|^\beta \leq \mathrm{g}_\beta\mathbf{x}^\beta$ for each $\beta \in \mathrm{Z}$ since $\mathrm{g}_\beta < 0$ due to (3.23). Hence, it holds $g(\mathbf{x}) \geq g(|\mathbf{x}|)$. Thus, $g \geq 0$ on $\mathbb{R}^{\mathrm{n}}$ if and only if $g \geq 0$ on $\mathbb{R}^{\mathrm{n}}_+$.

3.) Next, let $\mathrm{C} \subseteq \mathbb{Z}^{\mathrm{n}}$ be a set of vectors such that $(\mathrm{S}\backslash\{\mathbf{0}\}) \cup \mathrm{C}$ forms a basis of $\mathbb{R}^{\mathrm{n}}$. This is possible since the elements of S are affinely independent and $\mathbf{0} \in \mathrm{S}$ by (A3). Let $\mathbf{T}$ be a matrix, whose columns are the exponents in $(\mathrm{S}\backslash\{\mathbf{0}\}) \cup \mathrm{C}$. The matrix $\mathbf{T}$ is invertible by construction. Hence, it follows from Cramer's rule [61, Thm. 2.9] that $\det(\mathbf{T})\mathbf{T}^{-1} \in \mathbb{Z}^{\mathrm{n}\times\mathrm{n}}$ is an integer matrix. Let $\bar{\mathbf{T}} := 2|\det(\mathbf{T})|\mathbf{T}^{-1}$. Then for each $\alpha \in \mathrm{S}\backslash\{\mathbf{0}\}$ there exists $\mathrm{i} \in [\mathrm{n}]$ such that

$$\bar{\mathbf{T}}\alpha = 2|\det(\mathbf{T})|\mathrm{e}^{\mathrm{i}} \in 2\mathbb{N}^{\mathrm{n}}.$$

The set $\bar{\mathrm{S}} := \{\bar{\mathbf{T}}\alpha : \alpha \in \mathrm{S}\}$ satisfies (3.24), its elements are affinely independent and $\mathbf{0} \in \bar{\mathrm{S}}$. Furthermore, $\bar{\mathrm{Z}} := \{\bar{\mathbf{T}}\beta : \beta \in \mathrm{Z}\}$ is a subset of relint(conv($\bar{\mathrm{S}}$)) because of

$$\bar{\mathbf{T}}\beta = \sum_{\alpha\in\mathrm{S}} \mu_\alpha^\beta \bar{\mathbf{T}}\alpha$$

for $\mathrm{CR}(\mathrm{S}, \mu^\beta) \in \mathcal{F}$. Let $\bar{g} := \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha x^{\bar{\mathbf{T}}\alpha} \geq 0$. We claim that

$$g \geq 0 \text{ on } \mathbb{R}^{\mathrm{n}}_{++} \iff \bar{g} \geq 0 \text{ on } \mathbb{R}^{\mathrm{n}}_{++}. \tag{3.25}$$

[7] Recall that $\mathrm{g}_\alpha = \mathrm{f}_\alpha$ for $\alpha \in \mathrm{A}\backslash\{\mathbf{0}\}$.
[8] Recall that $\#\mathrm{Z} \geq 1$.

Since $\mathbb{R}^{\mathrm{n}}_{++}$ is dense in $\mathbb{R}^{\mathrm{n}}_{+}$, (3.25) is equivalent to

$$g \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}}_{+} \iff \bar{g} \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}}_{+},$$

which in turn concludes the third part of the proof. For the claim (3.25) observe that for all $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}_{++}$ there exists $\mathbf{w} \in \mathbb{R}^{\mathrm{n}}$ such that $\mathrm{x_i} = \exp(\mathrm{w_i})$ for all $\mathrm{i} \in [\mathrm{n}]$. Hence,

$$g \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}}_{++} \iff g((\exp(\mathrm{w_i}))_{\mathrm{i}\in[\mathrm{n}]}) \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}}.$$

The matrix $\bar{\mathbf{T}}$ is invertible and so is $\bar{\mathbf{T}}^\top$. Thus, for every $\mathbf{w} \in \mathbb{R}^{\mathrm{n}}$ there exists a unique vector $\mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ such that $\mathbf{w} = \bar{\mathbf{T}}^\top\mathbf{u}$. It follows from

$$\begin{aligned} g((\exp(\mathrm{w_i}))_{\mathrm{i}\in[\mathrm{n}]}) &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \prod_{\mathrm{i}\in[\mathrm{n}]} \exp(\mathrm{w_i})^{\alpha_\mathrm{i}} \\ &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \exp\Big(\sum_{\mathrm{i}\in[\mathrm{n}]} \mathrm{w_i}\alpha_\mathrm{i}\Big) \\ &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \exp\Big(\sum_{\mathrm{i}\in[\mathrm{n}]} (\bar{\mathbf{T}}^\top\mathbf{u})_\mathrm{i}\alpha_\mathrm{i}\Big) \\ &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \exp\Big(\sum_{\mathrm{i,j}\in[\mathrm{n}]} \bar{\mathbf{T}}_{\mathrm{j,i}}\mathrm{u_j}\alpha_\mathrm{i}\Big) \\ &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \exp\Big(\sum_{\mathrm{j}\in[\mathrm{n}]} (\bar{\mathbf{T}}\alpha)_\mathrm{j}\mathrm{u_j}\Big) \\ &= \sum_{\alpha\in\mathrm{A}} \mathrm{g}_\alpha \prod_{\mathrm{j}\in[\mathrm{n}]} \exp(\mathrm{u_j})^{(\bar{\mathbf{T}}\alpha)_\mathrm{j}} \\ &= \bar{g}((\exp(\mathrm{u_j}))_{\mathrm{j}\in[\mathrm{n}]}) \end{aligned}$$

that

$$\begin{aligned} g((\exp(\mathrm{w_i}))_{\mathrm{i}\in[\mathrm{n}]}) \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}} &\iff \bar{g}((\exp(\mathrm{u_i}))_{\mathrm{i}\in[\mathrm{n}]}) \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}} \\ &\iff \bar{g} \geq 0 \;\text{ on }\; \mathbb{R}^{\mathrm{n}}_{++}. \end{aligned}$$

This shows that we can assume that g and S satisfy (3.23) and (3.24) additionally to (A1) to (A8). Since $f = g + \lambda$, we have[9]

$$f = \sum_{\alpha\in\mathrm{S}} \mathrm{f}_\alpha x^\alpha + \sum_{\beta\in\mathrm{Z}} \mathrm{f}_\beta x^\beta \;\text{ with }\; \mathbf{f}_{\mathrm{S}\setminus\{\mathbf{0}\}} > \mathbf{0} \;\text{ and }\; \mathbf{f}_\mathrm{Z} < 0. \tag{3.26}$$

[9]Recall that $\mathrm{f}_\alpha = \mathrm{g}_\alpha$ for all $\alpha \in \mathrm{A}\setminus\{\mathbf{0}\}$.

4.) From (3.26) and (3.24) it follows that f is coercive, see [10, Prop. 2.37] for a proof. That means that the lower level sets $\mathrm{L}_\mathrm{a} := \{\mathbf{x} \in \mathbb{R}^\mathrm{n} : f(\mathbf{x}) \leq \mathrm{a}\}$ are bounded for all $\mathrm{a} \in \mathbb{R}$. Let $\bar{\mathbf{x}}$ be an arbitrary point in $\mathbb{R}^\mathrm{n}_+$ and $\mathrm{B} \subseteq \mathbb{R}^\mathrm{n}$ be a full-dimensional compact ball that is centered in the origin and contains $\mathrm{L}_{f(\bar{\mathbf{x}})}$. Then

$$\begin{aligned} \mathrm{f}^{\mathrm{pop}} &= \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^\mathrm{n}_+ \cap \mathrm{L}_{f(\bar{\mathbf{x}})}\} \\ &= \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}^\mathrm{n}_+ \cap \mathrm{B}\}. \end{aligned}$$

The feasible set $\mathbb{R}^\mathrm{n}_+ \cap \mathrm{B}$ is compact and nonempty. Thus, there exists $\mathbf{x}^\star \in \mathbb{R}^\mathrm{n}_+$ with $\mathrm{f}^{\mathrm{pop}} = f(\mathbf{x}^\star)$, which concludes the fourth part.

5.) Lemma 3.18 ensures that the optimal value $\mathrm{f}^{\mathrm{sonc}}$ of (D-SONC) is attained. This in particular implies that there exist $\lambda \in \mathbb{R}$ and circuit polynomials $f^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathcal{P}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ such that

$$f - \lambda = \sum_{\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}} f^{\mathrm{CR}(\mathrm{S},\mu^\beta)}. \tag{3.27}$$

Recall that each $\beta \in \mathrm{Z}$ is unique to a circuit in $\mathcal{F}$. Thus, every circuit polynomial $f^{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ participating in the sum (3.27) must satisfy $\mathrm{f}_\beta^{\mathrm{CR}(\mathrm{S},\mu^\beta)} < 0$ for $\beta \in \mathrm{Z}$. Replacing g in the second part of this proof by $f^{\mathrm{CR}(\mathrm{S},\mu^\beta)}$ shows that for the circuit polynomials in the sum (3.27) it also holds

$$f^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathcal{P}(\mathbb{R}^\mathrm{n})_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \iff f^{\mathrm{CR}(\mathrm{S},\mu^\beta)} \in \mathcal{P}(\mathbb{R}^\mathrm{n}_+)_{\mathrm{CR}(\mathrm{S},\mu^\beta)}.$$

This implies that the feasible sets of

$$\begin{aligned} \text{maximize} \quad & \lambda \\ \text{for} \quad & \lambda \in \mathbb{R} \\ \text{subject to} \quad & f - \lambda \in \sum_{\mathrm{CR}(\mathrm{S},\mu^\beta) \in \mathcal{F}} \mathcal{P}(\mathbb{R}^\mathrm{n}_+)_{\mathrm{CR}(\mathrm{S},\mu^\beta)} \end{aligned} \tag{3.28}$$

and of (D-SONC)[10] coincide and so do their optimal values.

From Theorem 2.10 it follows that the optimal value $\mathrm{f}^{\mathrm{sonc}}$ of (3.28) can be computed

[10]Recall that (D-SONC) is equivalent to (D-RLX) for families of circuit patterns $\mathcal{F}$ and $\mathrm{K} = \mathbb{R}^\mathrm{n}$.

by its dual program, which has, due to (3.19), the form

$$\begin{aligned} \text{minimize} \quad & \langle \mathbf{f}_{\mathrm{A}}, \mathbf{v}_{\mathrm{A}} \rangle \\ \text{for} \quad & \mathbf{v} \in \mathbb{R}_{+}^{\mathrm{A}} \\ \text{subject to} \quad & \mathrm{v}_{\beta} \leq \prod_{\alpha \in \mathrm{S}} \mathrm{v}_{\alpha}^{\mu_{\alpha}} \quad \text{for all } \mathrm{CR}(\mathrm{S}, \mu^{\beta}) \in \mathcal{F}, \\ & \mathrm{v}_{\mathbf{0}} = 1. \end{aligned} \tag{3.29}$$

By assumption (3.24) there exists an injective mapping that maps $\alpha \in \mathrm{S}\backslash\{\mathbf{0}\}$ to $\mathrm{i} \in [\mathrm{n}]$, which is determined by $\alpha = \alpha_{\mathrm{i}}\mathbf{e}^{\mathrm{i}}$. To emphasize this correspondence, we write $\alpha(\mathrm{i})$ instead of α. Furthermore, because $\mathbf{f}_{\mathrm{Z}} < \mathbf{0}$, the objective function of (3.29) is strictly decreasing in v_{β} for each $\beta \in \mathrm{Z}$. Since each $\beta \in \mathrm{Z}$ is unique to a circuit in $\mathcal{F}$, any feasible solution[11] $\mathbf{v}$ of (3.29), such that there exists $\beta \in \mathrm{Z}$ with $\mathrm{v}_{\beta} < \prod_{\alpha(\mathrm{i}) \in \mathrm{S}} \mathrm{v}_{\alpha}^{\mu_{\alpha(\mathrm{i})}}$, can be improved by increasing v_{β} until

$$\mathrm{v}_{\beta} = \prod_{\alpha(\mathrm{i}) \in \mathrm{S}} \mathrm{v}_{\alpha(\mathrm{i})}^{\mu_{\alpha(\mathrm{i})}} \tag{3.30}$$

holds. Thus, we replace the inequality constraint in (3.29) by (3.30). Due to $\mathbf{x}^{\alpha(\mathrm{i})} = \mathrm{x}_{\mathrm{i}}^{\alpha(\mathrm{i})_{\mathrm{i}}}$, for any feasible solution $\mathbf{v}$ there exists $\mathbf{x} \in \mathbb{R}_{+}^{\mathrm{n}}$ that solves

$$\mathrm{v}_{\alpha(\mathrm{i})} = \mathrm{x}_{\mathrm{i}}^{\alpha(\mathrm{i})_{\mathrm{i}}} \qquad \text{for all } \alpha(\mathrm{i}) \in \mathrm{S}\backslash\{\mathbf{0}\}.$$

Because of the constraint (3.30) the solution $\mathbf{x}$ satisfies

$$\mathrm{v}_{\beta} = \prod_{\alpha(\mathrm{i}) \in \mathrm{S}} \left(\mathrm{v}_{\alpha(\mathrm{i})}\right)^{\mu^{\beta}_{\alpha(\mathrm{i})}} = \prod_{\alpha(\mathrm{i}) \in \mathrm{S}} (\mathrm{x}_{\mathrm{i}}^{\alpha(\mathrm{i})_{\mathrm{i}}})^{\mu^{\beta}_{\alpha(\mathrm{i})}} = \mathbf{x}^{\beta} \qquad \text{for all } \beta \in \mathrm{Z}.$$

Thus, $\mathbf{v} = \mathrm{m}(\mathbf{x})_{\mathrm{A}}$. It follows that (3.29) coincides with

$$\inf\{\langle \mathbf{f}, \mathrm{m}(\mathbf{x})_{\mathrm{A}} \rangle : \mathbf{x} \in \mathbb{R}_{+}^{\mathrm{n}}\} = \inf\{f(\mathbf{x}) : \mathbf{x} \in \mathbb{R}_{+}^{\mathrm{n}}\} = \mathrm{f}^{\mathrm{pop}},$$

which shows $\mathrm{f}^{\mathrm{sonc}} = \mathrm{f}^{\mathrm{pop}}$. □

[11] A feasible solution exists, take for example $\mathbf{e}^{\mathbf{0},\mathrm{A}}$.

Using the optimal solution $\mathbf{x}^\star$ of (POP) from step 4 we can compute an optimal solution $\mathbf{v}^\star$ of (3.29) by setting

$$\begin{aligned} \mathrm{v}^\star_{\alpha(\mathrm{i})} &:= (\mathrm{x}^\star_\mathrm{i})^{\alpha(\mathrm{i})_\mathrm{i}} && \text{for all } \alpha(\mathrm{i}) \in \mathrm{S}\backslash\{\mathbf{0}\}, \\ \mathrm{v}^\star_\beta &:= (\mathbf{x}^\star)^\beta && \text{for all } \beta \in \mathrm{Z}. \end{aligned}$$

3.7 Scaled-Diagonally-Dominant Sums of Squares

Another relatively new approach uses scaled-diagonally-dominant sums of squares (SDSOS) [1], that is a Positivstellensatz based on SOS polynomials with restricted monomial support. These supports correspond to the following pattern type:

Definition 3.20 (Binomial Square Pattern).
Let $\alpha, \beta \in \mathbb{N}^\mathrm{n}$ with $\alpha \neq \beta$. We define the binomial square pattern (BS) as

$$\mathrm{BS}(\alpha, \beta) := \{2\alpha, 2\beta, \alpha + \beta\}.$$

Furthermore, let $\mathcal{F}$ be a family of binomial square patterns. Modifying the in [1] established notation we refer to a polynomial f that satisfies

$$f \in \sum_{\mathrm{P} \in \mathcal{F}} \mathcal{P}(\mathbb{R}^\mathrm{n})_\mathrm{P} \tag{3.31}$$

as scaled-diagonally-dominant sums of squares (SDSOS) and the Minkowski sum of cones as SDSOS cone with respect to $\mathcal{F}$.[12] $\triangle$

As with the SONC cone this slightly extends the original definition of the SDSOS cone by the family $\mathcal{F}$. One can see that a binomial square pattern is a special circuit pattern:

$$\mathrm{BS}(\alpha, \beta) = \mathrm{CR}(\{2\alpha, 2\beta\}, (1/2, 1/2)). \tag{3.32}$$

Like SONC, SDSOS are usually employed for unconstrained problems.

In the following we show that our notion of SDSOS in terms of polynomial cones matches the definition given in [1, Def. 3.2].

[12] As before, whenever we write $\mathrm{BS}(\mathrm{S}, \mu^\beta)$ we imply that the parameters α and β satisfy the assumptions of Definition 3.20.

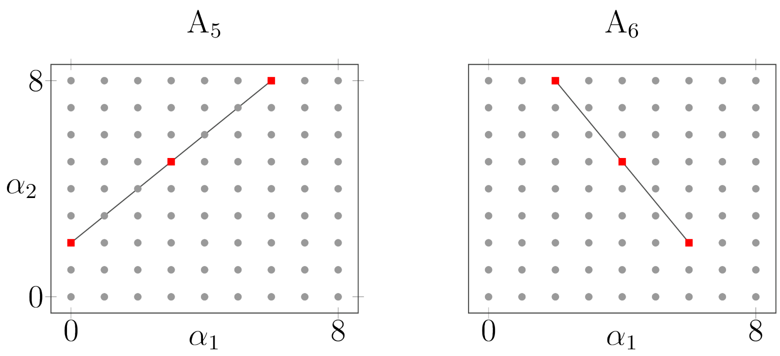

Figure 3.8: Visualization of two binomial square patterns as described in Plot Set Up 3.2. **Left:** $\mathrm{BS}((0,1),(3,4)) = \{(0,2),(3,5),(6,8)\}$. **Right:** $\mathrm{BS}((1,4),(3,1)) = \{(2,8),(4,5),(6,2)\}$.

Lemma 3.21.
Polynomials that are nonnegative on $\mathbb{R}^n$ with monomial support in $\mathrm{BS}(\alpha,\beta)$ are SOS. △

Proof. We define $\mu := (1/2, 1/2) \in \mathbb{R}^{\{2\alpha,2\beta\}}$. First, let $\alpha+\beta \notin 2\mathbb{N}^n$. From (3.32) and Corollary 3.17 follows that

$$f \in \mathcal{P}(\mathbb{R}^n)_{\mathrm{BS}(\alpha,\beta)} \iff \mathbf{f}_{\{2\alpha,2\beta\}} \geq \mathbf{0} \text{ and } 4\mathrm{f}_{2\alpha}\mathrm{f}_{2\beta} - (\mathrm{f}_{\alpha+\beta})^2 \geq 0. \tag{3.33}$$

Comparing the determinants of the principal minors of

$$\mathbf{F} := \begin{pmatrix} \mathrm{f}_{2\alpha} & \frac{\mathrm{f}_{\alpha+\beta}}{2} \\ \frac{\mathrm{f}_{\alpha+\beta}}{2} & \mathrm{f}_{2\beta} \end{pmatrix} \tag{3.34}$$

we see that (3.33) is equivalent to $\mathbf{F} \in \mathbb{S}_+^{\{\alpha,\beta\}}$. Hence, $f \in \mathcal{P}(\mathbb{R}^n)_{\mathrm{BS}(\alpha,\beta)}$ if and only if

$$f = \mathrm{f}_{2\alpha}x^{2\alpha} + \mathrm{f}_{\alpha+\beta}x^{\alpha+\beta} + \mathrm{f}_{2\beta}x^{2\beta} = \mathrm{m}(x)_{\{\alpha,\beta\}}^\top \mathbf{F}\,\mathrm{m}(x)_{\{\alpha,\beta\}}.$$

The above shows that $f \in \mathrm{SOS}(\mathrm{BS}(\alpha,\beta))$, which shows the assertion for $\alpha+\beta \notin 2\mathbb{N}^n$. Next, for $\alpha+\beta \in 2\mathbb{N}^n$ it follows from (3.32) and Corollary 3.17 that $f \in \mathcal{P}(\mathbb{R}^n)_{\mathrm{BS}(\alpha,\beta)}$

$$\begin{aligned} f \in \mathcal{P}(\mathbb{R}^n)_{\mathrm{BS}(\alpha,\beta)} \iff & \mathbf{f}_{\{2\alpha,2\beta\}} \geq \mathbf{0},\ 4\mathrm{f}_{2\alpha}\mathrm{f}_{2\beta} - \mathrm{t}^2 \geq 0 \text{ and} \\ & \mathrm{f}_{\alpha+\beta} = \mathrm{t} + \mathrm{u} \text{ with } \mathrm{u} \in \mathbb{R}_+. \end{aligned} \tag{3.35}$$

Thus, (3.35) is equivalent to

$$\mathbf{f} = \left(\mathrm{f}_{\{2\alpha,2\beta\}}, \mathrm{t}+\mathrm{u}\right) \text{ with } \bar{\mathbf{F}} := \begin{pmatrix} \mathrm{f}_{2\alpha} & \frac{\mathrm{t}}{2} \\ \frac{\mathrm{t}}{2} & \mathrm{f}_{2\beta} \end{pmatrix} \in \mathbb{S}_{+}^{\{\alpha,\beta\}} \text{ and } \mathrm{u} \in \mathbb{R}_{+}. \tag{3.36}$$

Hence, $f \in \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{BS}(\alpha,\beta)}$ if and only if $f = g + \mathrm{u}x^{\alpha+\beta}$, where

$$g = \mathrm{m}(x)_{\{\alpha,\beta\}}^{\top} \bar{\mathbf{F}}\, \mathrm{m}(x)_{\{\alpha,\beta\}} \in \mathrm{SOS}(\mathrm{BS}(\alpha,\beta)) \text{ and } \mathrm{u} \geq 0. \tag{3.37}$$

Observe that $\mathrm{u}x^{\alpha+\beta}$ is SOS since $\alpha+\beta \in 2\mathbb{N}^{\mathrm{n}}$. Thus, g can absorb $\mathrm{u}x^{\alpha+\beta}$ into its squares and (3.37), which shows the assertion for $\alpha+\beta \in 2\mathbb{N}^{\mathrm{n}}$. □

The proof of Lemma 3.21 shows that we can characterize SDSOS cones in the following way: Let $\mathcal{F}$ be a family of binomial square patterns and $\mathcal{F}^{\mathrm{e}} := \{\mathrm{BS}(\alpha,\beta) \in \mathcal{F} : \alpha+\beta \in 2\mathbb{N}^{\mathrm{n}}\}$. Then $f \in \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{BS}(\alpha,\beta)}$ if and only if there exist matrices $\mathbf{F}(\alpha,\beta) \in \mathbb{S}_{+}^{\{\alpha,\beta\}}$ for all $\mathrm{BS}(\alpha,\beta) \in \mathcal{F}$ and $\mathrm{u}_{\alpha+\beta} \in \mathbb{R}_{+}$ for all $\mathrm{BS}(\alpha,\beta) \in \mathcal{F}^{\mathrm{e}}$ such that

$$f = \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathrm{m}(x)_{\{\alpha,\beta\}}^{\top} \mathbf{F}(\alpha,\beta)\, \mathrm{m}(x)_{\{\alpha,\beta\}} + \sum_{\mathrm{BS}(\omega,\gamma)\in\mathcal{F}^{\mathrm{e}}} \mathrm{u}_{\omega+\gamma} x^{\omega+\gamma}. \tag{3.38}$$

Corollary 3.22 ([1, Thm. 3.6]).
A polynomial $f \in \mathbb{R}[x]_{\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}}$ is SDSOS if and only if it can be written as

$$f = \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathrm{m}(x)_{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}}^{\top} \mathbf{M}(\alpha,\beta)\, \mathrm{m}(x)_{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}}, \tag{3.39}$$

where $\mathcal{F}$ contains all binomial square patterns in $\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}$, i.e.

$$\mathcal{F} = \left\{ \mathrm{BS}(\alpha,\beta) : \alpha,\beta \in \mathbb{N}^{\mathrm{n}}_{\mathrm{d}} \text{ with } \alpha \neq \beta \right\},$$

and $\mathbf{M}(\alpha,\beta) \in \mathbb{R}^{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}} \times \mathbb{N}^{\mathrm{n}}_{\mathrm{d}}}$ for $\alpha,\beta \in \mathbb{N}^{\mathrm{n}}_{\mathrm{d}}$ with $\alpha \neq \beta$ are PSD matrices having at most the 4 nonzero entries: $\mathbf{M}(\alpha,\beta)_{\alpha,\alpha}, \mathbf{M}(\alpha,\beta)_{\alpha,\beta}, \mathbf{M}(\alpha,\beta)_{\beta,\alpha}$ and $\mathbf{M}(\alpha,\beta)_{\beta,\beta}$. △

Proof. First observe that all binomial square patterns that can contribute to the decomposition of f are elements of $\mathcal{F}$. Thus, f is a SDSOS if and only if

$$f \in \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{BS}(\alpha,\beta)}.$$

Using (3.38) and $\mathcal{F}^{\mathrm{e}} := \{\mathrm{BS}(\alpha,\beta) \in \mathcal{F} : \alpha + \beta \in 2\mathbb{N}^{\mathrm{n}}\}$ this is equivalent to the existence of matrices $\mathbf{F}(\alpha,\beta) \in \mathbb{S}_{+}^{\{\alpha,\beta\}}$ for all $\mathrm{BS}(\alpha,\beta) \in \mathcal{F}$ and scalars $\mathrm{u}_{\omega+\gamma} \in \mathbb{R}_{+}$ for all $\mathrm{BS}(\omega,\gamma) \in \mathcal{F}^{\mathrm{e}}$ such that

$$f = \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathrm{m}(x)_{\{\alpha,\beta\}}^{\top} \mathbf{F}(\alpha,\beta)\, \mathrm{m}(x)_{\{\alpha,\beta\}} + \sum_{\mathrm{BS}(\omega,\gamma)\in\mathcal{F}^{\mathrm{e}}} \mathrm{u}_{\omega+\gamma} x^{\omega+\gamma}. \tag{3.40}$$

We express $\mathrm{u}_{\omega+\gamma} \in \mathbb{R}_{+}$ as

$$\tilde{\mathbf{F}}\,(\omega,\gamma) := \begin{pmatrix} 0 & 0 \\ 0 & \mathrm{u}_{\omega+\gamma} \end{pmatrix} \in \mathbb{S}_{+}^{\{\mathbf{0},(\omega+\gamma)/2\}}.$$

Because of $\omega+\gamma \in 2\mathbb{N}^{\mathrm{n}}$ for all $\mathrm{BS}(\omega,\gamma) \in \mathcal{F}^{\mathrm{e}}$, we have $(\omega+\gamma)/2 \in \mathbb{N}_{\mathrm{d}}^{\mathrm{n}}$. With that we can write (3.40) as

$$\begin{aligned} f = &\sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathrm{m}(x)_{\{\alpha,\beta\}}^{\top} \mathbf{F}(\alpha,\beta)\, \mathrm{m}(x)_{\{\alpha,\beta\}} + \\ &\sum_{\mathrm{BS}(\omega,\gamma)\in\mathcal{F}^{\mathrm{e}}} \mathrm{m}(x)_{\{\mathbf{0},(\omega+\gamma)/2\}}^{\top} \tilde{\mathbf{F}}\,(\omega,\gamma)\, \mathrm{m}(x)_{\{\mathbf{0},(\omega+\gamma)/2\}}. \end{aligned}$$

Observe that $\mathrm{BS}(\mathbf{0}, (\omega+\gamma)/2) \in \mathcal{F}$. We define new matrices $\tilde{\mathbf{M}}(\alpha,\beta)$ for $\mathrm{BS}(\alpha,\beta) \in \mathcal{F}$ by

$$\tilde{\mathbf{M}}(\mathbf{0}, (\omega+\gamma)/2) := \mathbf{F}\,(\mathbf{0}, (\omega+\gamma)/2) + \tilde{\mathbf{F}}\,(\omega,\gamma)$$

for all $\mathrm{BS}(\omega,\gamma) \in \mathcal{F}^{\mathrm{e}}$ and otherwise

$$\tilde{\mathbf{M}}(\alpha,\beta) := \mathbf{F}(\alpha,\beta).$$

These matrices satisfy $\tilde{\mathbf{M}}(\alpha,\beta) \in \mathbb{S}_{+}^{\{\alpha,\beta\}}$ and $f = \sum_{\mathrm{BS}(\alpha,\beta)\in\mathcal{F}} \mathrm{m}(x)_{\{\alpha,\beta\}}^{\top} \tilde{\mathbf{M}}(\alpha,\beta)\, \mathrm{m}(x)_{\{\alpha,\beta\}}$. Finally, embedding $\tilde{\mathbf{M}}(\alpha,\beta)$ in $\mathbb{R}^{\mathbb{N}_{\mathrm{d}}^{\mathrm{n}}\times\mathbb{N}_{\mathrm{d}}^{\mathrm{n}}}$ using

$$\mathbf{M}(\alpha,\beta)_{\eta,\xi} := \begin{cases} \tilde{\mathbf{M}}(\alpha,\beta)_{\eta,\xi} & \text{if } \eta,\xi \in \{\alpha,\beta\}, \\ 0 & \text{else} \end{cases} \qquad \text{for all } \eta,\xi \in \mathbb{N}_{\mathrm{d}}^{\mathrm{n}}$$

yields the assertion. □

Corollary 3.22 shows that Definition 3.20 harmonizes with the definition of SDSOS in [1, Def. 3.3]. It is well known that $\mathbb{S}_+^{\{\alpha,\beta\}}$ is linear isomorphic to the three dimensional second-order cone (SOC), see for example [8]. Since there exist efficiently computable barrier functions for SOC the current implementation of SDSOS successfully uses SOC to solve (D-RLX) for families $\mathcal{F}$ of binomial square patterns [1]. We note that the authors of [1] also provide a linear programming (LP) approximation of the SDSOS cone, which is called diagonally-dominant sums of squares (DSOS).

Chapter 4

Truncated Submonoids

In this chapter we introduce new pattern types. These pattern types will be derived from a basic pattern type called truncated submonoid. While the pattern types discussed up to now were rather static in terms of their size, truncated submonoids are adaptable with respect to this matter.

As in the previous chapter we illustrate the patterns using different exponent sets.

Plot Set Up 4.1.
Among the exponent sets, that we use for the illustration of the new pattern are the sets from Plot Set Up 3.2 and

$$\begin{aligned}
\mathrm{A}_7 &:= \{(6,2),(6,3),(6,4),(6,5),(6,6),(7,3),(7,4),(7,5),(7,6),\\
&\qquad (8,4),(8,5),(8,6),(9,5),(9,6),(10,6)\},\\
\mathrm{A}_8 &:= \{(3,0),(3,1),(3,2),(3,3),(3,4),(3,5)\},\\
\mathrm{A}_9 &:= \{(0,3),(2,3),(5,3)\},\\
\mathrm{A}_{10} &:= \{(0,0),(2,2),(4,4),(6,6),(8,8)\},\\
\mathrm{A}_{11} &:= \{(2,8),(3,7),(4,6),(5,5),(6,4),(7,3),(8,2)\},\\
\mathrm{A}_{12} &:= \{(0,2),(2,3),(4,4),(6,5),(8,6)\},\\
\mathrm{A}_{13} &:= \{(2,2),(2,3),(2,4),(5,2),(5,3),(6,2)\},\\
\mathrm{A}_{14} &:= \{(4,0),(5,0),(5,3),(6,0),(6,3),(6,6)\},\\
\mathrm{A}_{15} &:= \{(2,2),(2,3),(2,4),(2,5),(2,6),(3,3),(3,4),(3,5),(3,6),\\
&\qquad (4,4),(4,5),(4,6),(5,5),(5,6),(6,6)\}. \qquad \triangle
\end{aligned}$$

Definition 4.2 (Truncated Submonoid).
Let $\mathrm{B} \subseteq \mathbb{N}^{\mathrm{n}}$ *be a nonempty finite set,* $\eta \in 2\mathbb{N}^{\mathrm{n}} \cap \mathrm{B}$ *be an exponent and* $\Gamma = (\gamma^1, \ldots, \gamma^{\mathrm{k}}) \in \mathbb{Z}^{\mathrm{n}\times\mathrm{k}}$ *with* $\mathrm{k} \in [\mathrm{n}]$ *be a matrix, whose columns* γ^{i} *are nonzero vectors with pairwise disjoint supports that satisfy* $\eta + \gamma^{\mathrm{i}} \in \mathrm{B}$*. We denote by* $\Lambda^+(\Gamma)$ *the*

submonoid of $(\mathbb{Z}^n, +, \mathbf{0})$ *generated by the columns of* Γ*, that is*

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

and call

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

the k*-variate truncated submonoid (TS). To avoid clumsy notation we overload the operator of truncated submonoids for* $\eta = \mathbf{0}$ *by* $\mathrm{TS}(\Gamma, \mathrm{B}) := \mathrm{TS}(\mathbf{0}, \Gamma, \mathrm{B})$.[1] △

Note that the vectors $\gamma^1, \ldots, \gamma^k$ are linearly independent. For an illustration of different truncated submonoids see Figure 4.1.

We typically use for the truncation sets like $\mathrm{B} = \eta \pm \mathbb{N}^n_r$ or $\mathrm{B} = \mathbb{N}^n_r$ for appropriate $r \in \mathbb{N}$. However, allowing other sets B as well, makes truncated submonoids more versatile. For example we can express the multilinear pattern $\mathrm{ML}(\alpha, \mathrm{I})$ as $\mathrm{TS}(\Gamma, \mathrm{B})$ by choosing appropriate generators from $\left\{ \sum_{i \in J} \alpha_i \mathbf{e}^i : J \subseteq [n] \text{ with } \#J \geq 1 \right\}$ and an appropriate set B. Furthermore, with $\Gamma = (\mathbf{e}^1, \ldots, \mathbf{e}^n)$ and $\mathrm{B} = \mathbb{N}^n_{2d}$ one arrives at the extreme case, where $\mathrm{TS}(\Gamma, \mathrm{B}) = \mathbb{N}^n_{2d}$ is the pattern of a moment relaxation, see for example the left subplot of Figure 4.1.

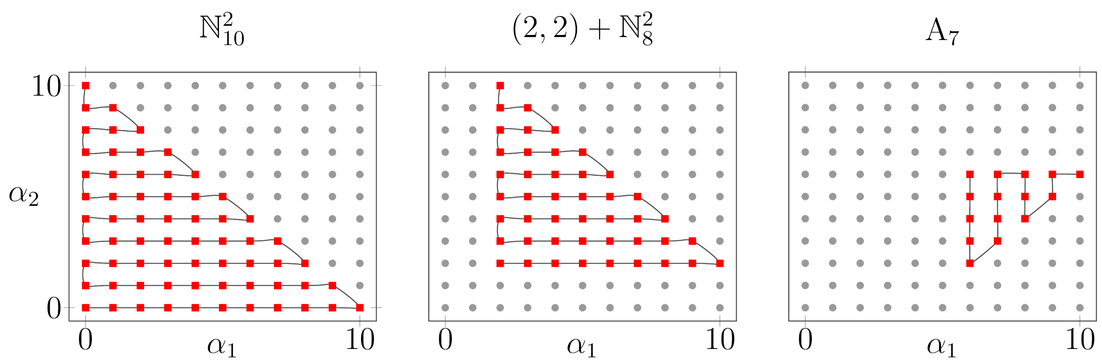

Figure 4.1: Visualization of possible truncated submonoids as described in Plot Set Up 3.2. **Left:** $\mathrm{TS}(\Gamma, \mathrm{B})$ for $\Gamma = (\mathbf{e}^1, \mathbf{e}^2), \mathrm{B} = \mathbb{N}^2_{10}$. **Middle:** $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ for $\eta = (2, 2)$, $\Gamma = (\mathbf{e}^1, \mathbf{e}^2), \mathrm{B} = \mathbb{N}^2_{12}$. **Right:** $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ for $\eta = (6, 6)$, $\Gamma = (\mathbf{e}^1, -\mathbf{e}^2), \mathrm{B} = \mathrm{A}_7$.

By slight abuse of notation we will interpret $\mathrm{m}(\mathbf{x})_\Gamma$ in the following as $(x^{\gamma_i})_{i \in [k]} \in \mathbb{R}^k$.

[1] From this point onwards, whenever we write $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ we imply that the parameters $k, \eta, \Gamma = (\gamma^1, \ldots, \gamma^k)$ and B satisfy the assumptions of Definition 4.2.

Proposition 4.3.
Let $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ *be a* k*-variate truncated submonoid and*

$$\mathrm{L} := \bigcup_{\mathrm{i} \in [\mathrm{n}]} \{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : \text{there exists } \mathrm{i} \in [\mathrm{n}] \text{ with } \mathrm{x}_{\mathrm{i}} = 0\}.$$

If $\mathrm{K} \backslash \mathrm{L}$ *is dense in* K*, then*

$$\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} = \Big\{ x^{\eta}\, \tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}} \Big\},$$

where $\tilde{\mathrm{K}} := \mathrm{cl}\{\mathrm{m}(\mathbf{x})_{\Gamma} : \mathbf{x} \in \mathrm{K} \backslash \mathrm{L}\} \subseteq \mathbb{R}^{\mathrm{k}}$ *and*

$$\tilde{\mathrm{P}} := \big\{ \omega \in \mathbb{N}^{\mathrm{k}} : \eta + \Gamma\omega \in \mathrm{TS}(\eta, \Gamma, \mathrm{B}) \big\}.$$

△

Proof. Since $\mathrm{K} \backslash \mathrm{L}$ is dense in K, a polynomial g is nonnegative on K if and only if g is nonnegative on $\mathrm{K} \backslash \mathrm{L}$. Since $\eta \in 2\mathbb{N}^{\mathrm{n}}$, $\mathbf{x}^{\eta} > 0$ for all $\mathbf{x} \in \mathrm{K} \backslash \mathrm{L}$. Moreover, $\mathbf{x}^{\gamma^{\mathrm{i}}}$ is well defined for all $\mathbf{x} \in \mathrm{K} \backslash \mathrm{L}$ and $\mathrm{i} \in [\mathrm{k}]$ – even when γ^{i} has negative entries. Thus,

$$\begin{aligned}
\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} &= \Big\{ g \in \mathbb{R}[x] : g = \sum_{\omega \in \tilde{\mathrm{P}}} \mathrm{g}_{\eta+\Gamma\omega} x^{\eta+\Gamma\omega} \geq 0 \text{ on } \mathrm{K} \backslash \mathrm{L} \Big\} \\
&= \Big\{ x^{\eta}\, h \in \mathbb{R}[x] : h = \sum_{\omega \in \tilde{\mathrm{P}}} \mathrm{h}_{\Gamma\omega} x^{\Gamma\omega} \geq 0 \text{ on } \mathrm{K} \backslash \mathrm{L} \Big\} \\
&= \Big\{ x^{\eta}\, \tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : \tilde{h} = \sum_{\omega \in \tilde{\mathrm{P}}} \tilde{\mathrm{h}}_{\omega} y^{\omega} \geq 0 \text{ on } \mathrm{m}(\mathrm{K} \backslash \mathrm{L})_{\Gamma} \Big\}.
\end{aligned}$$

The last inequality follows from substituting $y_{\mathrm{i}} := x^{\gamma_{\mathrm{i}}}$ for $\mathrm{i} \in [\mathrm{k}]$.[2] Due to the way we have defined $\tilde{\mathrm{K}}$, $\mathrm{m}(\mathrm{K} \backslash \mathrm{L})_{\Gamma}$ is dense in $\tilde{\mathrm{K}}$. It follows

$$\begin{aligned}
\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} &= \Big\{ x^{\eta}\, \tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : \tilde{h} = \sum_{\omega \in \tilde{\mathrm{P}}} \tilde{\mathrm{h}}_{\omega} y^{\omega} \geq 0 \text{ on } \tilde{\mathrm{K}} \Big\} \\
&= \Big\{ x^{\eta}\, \tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}} \Big\}.
\end{aligned}$$

□

The next corollary is the dual formulation of Proposition 4.3.

[2]Note that h is not necessarily a polynomial since $\Gamma \in \mathbb{Z}^{\mathrm{n} \times \mathrm{k}}$ allows for negative components of the exponents.

Corollary 4.4.
Let the setting be as in Proposition 4.3. Then $\mathcal{C}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} \cong \mathcal{C}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}}$. △

Proof. We abbreviate $\mathrm{G} := \mathrm{TS}(\eta, \Gamma, \mathrm{B})$. Furthermore, we define a linear map $\mathrm{ind}_{\eta,\Gamma} : \mathbb{R}^{\tilde{\mathrm{P}}} \to \mathbb{R}^{\mathrm{G}}$ that takes care of index changes by $\mathrm{ind}_{\eta,\Gamma}(\tilde{\mathbf{v}}) = \mathbf{v}$, where

$$\mathrm{v}_{\eta+\Gamma\omega} := \tilde{\mathrm{v}}_{\omega} \quad \text{for } \omega \in \tilde{\mathrm{P}}.$$

Applying Lemma 2.6 $(*)$ and Proposition 4.3 $(\star)$ yields

$$\begin{aligned}
\mathcal{C}(\mathrm{K})_{\mathrm{G}} &\overset{(*)}{=} \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{G}} : \langle \mathbf{v}, \mathrm{vec}(h)_{\mathrm{G}} \rangle \geq 0 \text{ for all } h \in \mathcal{P}(\mathrm{K})_{\mathrm{G}}\right\} \\
&\overset{(\star)}{=} \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{G}} : \langle \mathbf{v}, \mathrm{vec}(x^{\eta}\, \tilde{h}(\mathrm{m}(x)_{\Gamma}))_{\mathrm{G}} \rangle \geq 0 \text{ for all } \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}}\right\} \\
&= \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{G}} : \sum_{\omega \in \tilde{\mathrm{P}}} \mathrm{v}_{\eta+\Gamma\omega}\, \mathrm{vec}(\tilde{\mathrm{h}})_{\omega} \geq 0 \text{ for all } \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}}\right\} \\
&\overset{(*)}{=} \left\{\mathbf{v} \in \mathbb{R}^{\mathrm{G}} : \text{there exists } \tilde{\mathbf{v}} \in \mathcal{C}(\tilde{\mathrm{K}})_{\tilde{\mathrm{P}}} \text{ with } \mathbf{v} = \mathrm{ind}_{\eta,\Gamma}(\tilde{\mathbf{v}})\right\}.
\end{aligned}$$ □

The next proposition describes the well-known relationship between PSD and SOS cones.

Proposition 4.5 ([37, Ch. 2.1]).
The cone of n*-variate SOS polynomials of degree at most* 2d *is the image of the cone of semidefinite matrices of size* $\binom{\mathrm{n}+\mathrm{d}}{\mathrm{n}}$ *under a linear transformation. More precisely, one has*

$$\mathrm{SOS}(\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}) = \left\{\mathrm{m}(x)^{\top}_{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}} \mathbf{T}\, \mathrm{m}(x)_{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}} : \mathbf{T} \in \mathbb{S}^{\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}}_{+}\right\}.$$ △

Propositions 4.3 and 4.5 give rise to the next definition.

Definition 4.6 ($\mathbb{N}^{\mathbf{k}}_{\mathbf{2d}}$-Description).
Let $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ *be a* k*-variate truncated submonoid,* $\mathrm{d}, \mathrm{r} \geq 0$ *be integers,* $\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]} \subseteq \mathbb{R}[y]_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}$ *be a set of* k*-variate polynomials and* $\tilde{\mathrm{d}}_{\mathrm{i}} := \lfloor \deg(g^{\mathrm{i}})/2 \rfloor$. *Furthermore, let*

$$\begin{aligned}
\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta} := \Big\{ x^{\eta} h(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : h = s^0 + \sum_{\mathrm{i}\in[\mathrm{r}]} s^{\mathrm{i}} g^{\mathrm{i}} \text{ with} \\
s^0 \in \mathrm{SOS}\left(\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}\right) \text{ and } s^{\mathrm{i}} \in \mathrm{SOS}\left(\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}-2\tilde{\mathrm{d}}_{\mathrm{i}}}\right) \Big\}.
\end{aligned} \tag{SD}$$

We call (SD) *an approximate* $\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$*-description of* $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ *if*

$$\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta} \subseteq \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$$

and a $\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$*-description of* $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ *if*

$$\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta} = \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}. \qquad \triangle$$

The set (SD) is related to the notion of quadratic modules and truncated quadratic modules, compare [38, 46]. In fact, if Γ is an identity matrix and $\nu = \mathbf{0}$, then $\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta}$ is a 2d-th truncated quadratic module. Using Proposition 4.5 we can replace the conic variables $s^0, \ldots, s^{\mathrm{r}}$ via lifting by semidefinite matrix variables of sizes $\binom{\mathrm{n}+\mathrm{d}}{\mathrm{d}}$, $\binom{\mathrm{n}+\mathrm{d}-\tilde{\mathrm{d}}_1}{\mathrm{d}-\tilde{\mathrm{d}}_1}$, $\ldots$, $\binom{\mathrm{n}+\mathrm{d}-\tilde{\mathrm{d}}_\mathrm{r}}{\mathrm{d}-\tilde{\mathrm{d}}_\mathrm{r}}$, respectively. Thus, replacing $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ in (D-RLX) by $\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta}$ makes (D-RLX) accessible to SDP solvers.

Corollary 4.7.
Additional to setting of Proposition 4.3, let $\mathrm{B} = \{\eta + \Gamma\omega : \omega \in \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}\} \subseteq \mathbb{N}^{\mathrm{n}}$ *and* $g^1, \ldots, g^{\mathrm{r}} \in \mathbb{R}[y]_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}$ *be* k*-variate polynomials such that*

$$\tilde{\mathrm{K}} = \{\mathbf{y} \in \mathbb{R}^{\mathrm{k}} : g^{\mathrm{i}}(\mathbf{y}) \geq 0 \ \text{ for all } \ \mathrm{i} \in [\mathrm{r}]\}. \tag{4.1}$$

Then $\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta}$ *is an approximate* $\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$*-description of* $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$. $\triangle$

Proof. From Proposition 4.3 follows with $\tilde{\mathrm{P}} = \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$ that

$$\begin{aligned}
\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} &= \left\{x^\eta\, \tilde{h}(\mathrm{m}(x)_\Gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}\right\} \\
&\supseteq \Bigg\{x^\eta\, \tilde{h}(\mathrm{m}(x)_\Gamma) \in \mathbb{R}[x] : \tilde{h} = s^0 + \sum_{\mathrm{i}\in[\mathrm{r}]} s^{\mathrm{i}} g^{\mathrm{i}} \ \text{ with} \\
&\qquad\qquad s^0 \in \mathrm{SOS}\left(\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}\right) \ \text{ and } \ s^{\mathrm{i}} \in \mathrm{SOS}\left(\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}-2\tilde{\mathrm{d}}_\mathrm{i}}\right)\Bigg\} \\
&= \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta},
\end{aligned}$$

where $\tilde{\mathrm{d}}_\mathrm{i} := \lfloor \deg(g^{\mathrm{i}})/2 \rfloor$. $\square$

The requirement (4.1) on $\tilde{\mathrm{K}}$ is rather strong. However, if there do not exist polynomials $g^{\mathrm{i}} \in \mathbb{R}[y]_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}$ such that (4.1) holds, one can use a set $\bar{\mathrm{K}}$ with $\mathrm{K} \subseteq \bar{\mathrm{K}}$ such that this is the case. This way one can relax $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ by $\mathcal{P}(\bar{\mathrm{K}})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$, which in turn can be relaxed by an approximate $\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$-description.

With Corollary 4.7 we can relax (D-RLX) by replacing $f^{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ with

$f^{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{2\mathrm{d}})_{\Gamma,\eta}$. The relaxation of (D-RLX) obtained this way is a SDP. Naturally, the computability of those SDPs depends on the degree[3] of the set $\tilde{\mathrm{P}}$ and the number of submonoid generators $\gamma^1, \ldots, \gamma^{\mathrm{k}}$. For practical purposes, it is desirable to choose k to be a relatively small number. Furthermore, by appropriately choosing the length of the vectors $\gamma^1, \ldots, \gamma^{\mathrm{k}} \in \mathrm{B} \setminus \{\mathbf{0}\}$ in correspondence with the choice of the set B, we manipulate the degree of $\tilde{\mathrm{P}}$. This allows to control the tractability of the pattern relaxation when used with truncated submonoids.

We would like to stress that in practice it is usually infeasible to use SOS relaxations for the original problem (POP) due to the size of these relaxations, see [7] for a theoretical justification. In contrast, we believe that one can use SOS relaxations together with truncated submonoids in (D-RLX), since we can control the number and length of the generators and $\deg(\tilde{\mathrm{P}})$ and therefore the size of the required SOS/PSD cones.

Throughout the rest of this chapter we combine Proposition 4.3 and well-known Positivstellensätze to derive new pattern types. The first corollary that we use is a special case of Putinar's theorem [51], which we have seen before in Theorem 3.11. See for example [6] for a short proof.

Corollary 4.8.
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ with $\mathbf{l} < \mathbf{u}$ and $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$. Furthermore, let $p(\mathbf{x})$ be a polynomial satisfying $p(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathrm{K}$. Then there exists an integer $\mathrm{d} \in \mathbb{N}$ such that

$$p \in \mathrm{SOS}(\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}) + \sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{u_i} - x_{\mathrm{i}})(x_{\mathrm{i}} - \mathrm{l_i})\, \mathrm{SOS}(\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}-2}). \qquad \triangle$$

By applying the shifting procedure to truncated submonoids we generate a new pattern type that is suited for cases where K is an axis-parallel box.

Corollary 4.9.
Let the setting be as in Corollary 4.7, where we specify $\eta = 0$, $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ for $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ with $\mathbf{l} < \mathbf{u}$ and $g^{\mathrm{i}}(y) = (\overline{\mathbf{x}}_{\mathrm{K}}^{\gamma_{\mathrm{i}}} - y_{\mathrm{i}})(y_{\mathrm{i}} - \underline{\mathbf{x}}_{\mathrm{K}}^{\gamma_{\mathrm{i}}}) \in \mathbb{R}[y]_{\mathbb{N}^{\mathrm{k}}_2}$ for $\mathrm{i} \in [\mathrm{k}]$. Furthermore, let $\xi \in \mathbb{N}^{\mathrm{n}}$ be a shift vector such that $\mathrm{supp}(\xi) \cap \mathrm{supp}(\gamma^{\mathrm{i}}) = \emptyset$ for all columns γ^{i} of Γ. Then for every $f \in \mathcal{P}(\mathrm{K})_{\xi+\mathrm{TS}(\Gamma,\mathrm{B})}$ and $\varepsilon > 0$ there exists $\tilde{\mathrm{d}} \in \mathbb{N}$ such that

$$f + \varepsilon \in \left\{ x^{\xi} g : \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}} \text{ and } \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}} \right\}. \qquad \triangle$$

[3]Recall that the degree of $\tilde{\mathrm{P}}$ is defined as $\deg(\tilde{\mathrm{P}}) := \max\{|\omega_1| + \cdots + |\omega_{\mathrm{k}}| : \omega \in \tilde{\mathrm{P}}\}$.

Proof. Proposition 2.14 applied to $\mathcal{P}(\mathrm{K})_{\xi+\mathrm{TS}(\Gamma,\mathrm{B})}$ yields

$$\mathcal{P}(\mathrm{K})_{\xi+\mathrm{TS}(\Gamma,\mathrm{B})} = \left\{ x^{\xi} g : \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})} \text{ and } \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})} \right\}.$$

Hence, there exists g such that

$$f + \varepsilon = x^{\xi} g \tag{4.2}$$

as well as $\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})}$ and $\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})}$. Next, observe that

$$\begin{aligned} \tilde{\mathrm{K}} :=& \operatorname{cl}\{\mathrm{m}(\mathbf{x})_{\Gamma} : \mathbf{x} \in \mathrm{K}\backslash \mathrm{L}\} \\ =& \{\mathbf{y} \in \mathbb{R}^{\mathrm{k}} : g^{1}(\mathbf{y}) \geq 0, \ldots, g^{\mathrm{k}}(\mathbf{y}) \geq 0\}. \end{aligned}$$

Thus, Proposition 4.3 states for $\mathrm{TS}(\Gamma, \mathrm{B})$ and K that

$$\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})} = \left\{ \tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}}} \right\}.$$

Hence, there exist $\tilde{h}^1, \tilde{h}^2 \in \mathcal{P}(\tilde{\mathrm{K}})_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}$ such that $\tilde{h}^1(\mathrm{m}(x)_{\Gamma}) = \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g$ and $\tilde{h}^2(\mathrm{m}(x)_{\Gamma}) = \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g$. Furthermore, from $f \in \mathcal{P}(\mathrm{K})_{\xi+\mathrm{TS}(\Gamma,\mathrm{B})}$ follows that $f + \varepsilon > 0$ on K and therefore $\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g > 0$ and $\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g > 0$ on K. This in turn implies that $\tilde{h}^1, \tilde{h}^2 > 0$ on $\tilde{\mathrm{K}}$.

Note that Corollary 4.8 can be expressed in terms of $\mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}}$: Let $\tilde{h} \in \mathcal{P}(\tilde{\mathrm{K}})_{\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}}$ with $\tilde{h} > 0$ on $\tilde{\mathrm{K}}$. Then there exists $\tilde{\mathrm{d}}$ such that $\tilde{h}(\mathrm{m}(x)_{\Gamma}) \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}}$.

Thus, there exist $\tilde{\mathrm{d}}_1$ and $\tilde{\mathrm{d}}_2$ such that $\tilde{h}^1(\mathrm{m}(x)_{\Gamma}) \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}_1})_{\Gamma,\mathbf{0}}$ and $\tilde{h}^2(\mathrm{m}(x)_{\Gamma}) \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}_2})_{\Gamma,\mathbf{0}}$. With $\tilde{\mathrm{d}} := \max\{\tilde{\mathrm{d}}_1, \tilde{\mathrm{d}}_2\}$ follows that $\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}}$ and $\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, g \in \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}, \mathbb{N}^{\mathrm{k}}_{2\tilde{\mathrm{d}}})_{\Gamma,\mathbf{0}}$. □

Figure 4.2 shows examples of shifted truncated submonoids.

Proposition 4.10 ([9]).
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l}, \mathbf{u}$ *and* $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$*. Then the moment body* $\mathcal{M}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})}$ *can be represented as a* k*-variate moment body by*[4]

$$\mathcal{M}\big(\mathrm{K}\big)_{\mathrm{TS}(\Gamma,\mathrm{B})} \cong \mathcal{M}\big(\,\mathrm{Box}\,\big(\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}\,\big)\big)_{\tilde{\mathrm{P}}},$$

with

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

△

[4]Here we use, by slight abuse of notation, $\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma} := (\underline{\mathbf{x}}_{\mathrm{K}}^{\gamma^{\mathrm{i}}})_{\mathrm{i}\in[\mathrm{k}]}$ and $\overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma} := (\overline{\mathbf{x}}_{\mathrm{K}}^{\gamma^{\mathrm{i}}})_{\mathrm{i}\in[\mathrm{k}]}$.

Proof. The desired representation is obtained by taking the convex hull of the left and the right hand side of $\mathrm{m}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})} \cong \mathrm{m}(\mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}))_{\tilde{\mathrm{P}}}$. Note that the isomorphy is due to the different indexing of the vector spaces that $\mathrm{m}(\mathrm{K})_{\mathrm{TS}(\Gamma,\mathrm{B})} \subseteq \mathbb{R}^{\mathrm{TS}(\Gamma,\mathrm{B})}$ and $\mathrm{m}(\mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\Gamma}))_{\tilde{\mathrm{P}}} \subseteq \mathbb{R}^{\tilde{\mathrm{P}}}$ live in. □

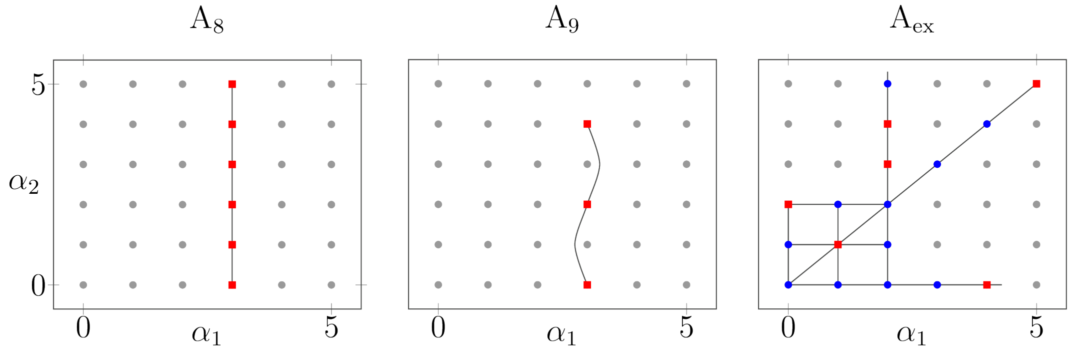

Figure 4.2: Visualization of possible shifted truncated submonoids as described in Plot Set Up 3.2. **Left:** $\xi^1 + \mathrm{TS}(\Gamma^1, \mathrm{B}^1)$ for $\xi^1 = (3,0)^\top, \Gamma^1 = \mathbf{e}^2, \mathrm{B}^1 = [5]_0^2$. **Middle:** $\xi^1 + \mathrm{TS}(\Gamma^2, \mathrm{B}^1)$ for $\Gamma^1 = 2\mathbf{e}^2$. **Right:** $\mathcal{F}^3 = \{\mathrm{TS}(\mathbf{e}^1, [4]_0^2), \mathbf{e}^2 + \mathrm{TS}(\mathbf{e}^1, [2]_0^2), 2\mathbf{e}^2 + \mathrm{TS}(\mathbf{e}^1, [2]_0^2), \mathrm{TS}(\mathbf{e}^2, [2]_0^2), \mathbf{e}^1 + \mathrm{TS}(\mathbf{e}^2, [2]_0^2), 2\mathbf{e}^1 + \mathrm{TS}(\mathbf{e}^1, [5]_0^2), \mathrm{TS}(\mathbf{1}, [5]_0^2)\}$.

Before we proceed with more specific patterns, we list the other Positivstellensätze that we will use for deriving new pattern types.

Theorem 4.11 (Hilbert, [40, Thm. 1.2.6] & [40, Prop. 1.2.1]).
It holds

$$\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathbb{N}_{\mathrm{d}}^{\mathrm{n}}} = \mathrm{SOS}(\mathbb{N}_{\mathrm{d}}^{\mathrm{n}})$$

if and only if one of the following holds

(H1) $\mathrm{n} = 1, \quad \mathrm{d} \in 2\mathbb{N},$

(H2) $\mathrm{n} \in \mathbb{N}, \quad \mathrm{d} = 2,$

(H3) $\mathrm{n} = 2, \quad \mathrm{d} = 4.$

Furthermore, in the case of $\mathrm{n} = 1$, $f \in \mathcal{P}(\mathbb{R})_{\mathbb{N}_{\mathrm{d}}}$ *if and only if* f *is a sum of at most two squared polynomials.* △

Corollary 4.12 (Pólya-Szegö [38, Thm. 3.21], Stieltjes [40, Cor. 3.1.3]).
Let $\mathrm{d} \in \mathbb{N}$. *Then*

$$\mathcal{P}(\mathbb{R}_+)_{[2\mathrm{d}]_0} = \mathrm{SOS}([2\mathrm{d}]_0) + x \cdot \mathrm{SOS}([2\mathrm{d}-2]_0).$$

△

Theorem 4.13 (Fekete [38, Thm. 3.23], Hausdorff [40, Cor. 3.1.5]).
Let d *be a nonnegative integer and* $\mathrm{l}, \mathrm{u} \in \mathbb{R}$ *with* $\mathrm{l} < \mathrm{u}$. *Then*

$$\mathcal{P}([\mathrm{l}, \mathrm{u}])_{[2\mathrm{d}]_0} = \mathrm{SOS}([2\mathrm{d}]_0) + (x - \mathrm{l})(\mathrm{u} - x) \cdot \mathrm{SOS}([2\mathrm{d} - 2]_0).$$

△

Theorem 4.14 (Švecov [40, Cor. 3.1.6]).
Let d *be a nonnegative integer and* $\mathrm{l}, \mathrm{u} \in \mathbb{R}$ *with* $\mathrm{l} < \mathrm{u}$. *Then*

$$\mathcal{P}((-\infty, \mathrm{l}] \cup [\mathrm{u}, \infty))_{[2\mathrm{d}]_0} = \mathrm{SOS}([2\mathrm{d}]_0) + (\mathrm{l} - x)(\mathrm{u} - x) \cdot \mathrm{SOS}([2\mathrm{d} - 2]_0).$$

△

Remark 4.15.
The reference for Theorem 4.14 does not state the degree bounds for the SOS polynomials, but just that

$$\mathcal{P}((-\infty, \mathrm{l}] \cup [\mathrm{u}, \infty))_{[2\mathrm{d}]_0} = \mathrm{SOS}(\mathbb{N}) + (\mathrm{l} - x)(\mathrm{u} - x) \cdot \mathrm{SOS}(\mathbb{N}).$$

However, the extreme rays of the cone $\mathcal{P}((-\infty, \mathrm{l}] \cup [\mathrm{u}, \infty))_{[2\mathrm{d}]_0}$ *are polynomials* g *that are nonnegative on* $(-\infty, \mathrm{l}] \cup [\mathrm{u}, \infty)$ *and have* $\deg(g)$ *many algebraic roots in* $(-\infty, \mathrm{l}] \cup [\mathrm{u}, \infty)$. *All such polynomials* g *are contained in* $\mathrm{SOS}([2\mathrm{d}]_0) \cup (\mathrm{l} - x)(\mathrm{u} - x) \cdot \mathrm{SOS}([2\mathrm{d} - 2]_0)$, *which shows the nontrivial inclusion of Theorem 4.14.* △

4.1 Chains

The next pattern type generalizes binomial square patterns.

Definition 4.16 (Chain).
Let $\eta \in 2\mathbb{N}^{\mathrm{n}}$, $\gamma \in \mathbb{Z}^{\mathrm{n}}$ *with* $\gamma \neq \mathbf{0}$ *and* $\mathrm{d} \in \mathbb{N}\backslash\{0\}$ *satisfy* $\eta + \mathrm{d}\gamma \geq \mathbf{0}$. *We call*

$$\mathrm{CH}(\eta, \gamma, \mathrm{d}) := \{\eta + \mathrm{i}\gamma : \mathrm{i} \in [\mathrm{d}]_0\}$$

a chain (CH).[5] *As with the truncated submonoids, we overload this notation for* $\eta = \mathbf{0}$ *by* $\mathrm{CH}(\gamma, \mathrm{d}) := \mathrm{CH}(\mathbf{0}, \gamma, \mathrm{d})$. △

Figure 4.3 demonstrates how the parameters d and γ in Definition 4.16 allow to control the size and orientation of chains.

[5]As usual, whenever we write $\mathrm{CH}(\eta, \gamma, \mathrm{d})$ we imply that the parameters η, γ and d satisfy the assumptions of Definition 4.16.

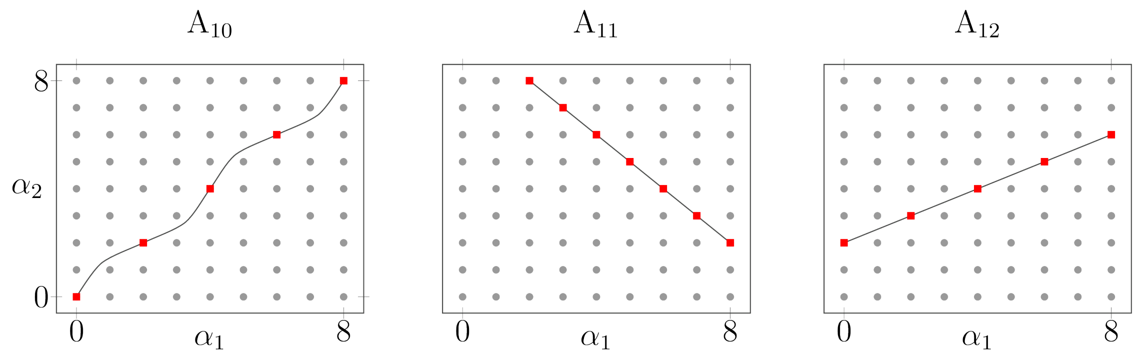

Figure 4.3: Visualization of possible chains as described in Plot Set Up 3.2. **Left:** $\mathrm{CH}((2,2),4)$. **Middle:** $\mathrm{CH}((2,8),(1,-1),6)$. **Right:** $\mathrm{CH}((0,2),(2,1),4)$.

Chains are univariate truncated submonoids since $\mathrm{CH}(\eta,\gamma,\mathrm{d}) = \mathrm{TS}(\eta,\gamma,\mathrm{B})$ with $\mathrm{B} = \{\eta + \mathrm{i}\gamma : \mathrm{i} \in [\mathrm{d}]_0\}$. Furthermore, a binomial square pattern is a special chain that contains only 3 exponents:

$$\begin{aligned}\mathrm{CH}(\eta,\gamma,2) &= \{\eta, \eta+\gamma, \eta+2\gamma\}\\ &= \mathrm{BS}(\eta/2, \eta/2+\gamma).\end{aligned}$$

Recall that for a binomial square pattern P and $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$ it is possible to enforce $f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P})}$ by PSD constraints and one 2×2 PSD matrix. The next corollaries show that $f^{\mathrm{CH}(\alpha,\gamma,\mathrm{k})} \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\alpha,\gamma,\mathrm{k})}$ can be enforced by semidefinite constraints of reasonable sizes, when $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$ or $\mathrm{K} = \mathrm{Box}(\mathbf{l},\mathbf{u})$ with $\mathbf{l} < \mathbf{u}$. Corollary 4.17 also yields an alternative way to show Lemma 3.21.

Corollary 4.17.
Let $\mathrm{CH}(\eta,\gamma,2\mathrm{d})$ be a chain. $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ has an $\mathrm{N}_{2\mathrm{d}}$-description. △

Proof. For $\gamma \notin 2\mathbb{N}^{\mathrm{n}}$ the assertion follows from Proposition 4.3 and (H1) Theorem 4.11. Then $\tilde{\mathrm{K}}$ and $\tilde{\mathrm{P}}$ of Proposition 4.3 are $\tilde{\mathrm{K}} = \mathbb{R}$ and $\tilde{\mathrm{P}} = [2\mathrm{d}]_0$. Hence,

$$\begin{aligned}\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} &= \left\{x^\eta \tilde{h}(x^\gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\mathbb{R})_{[2\mathrm{d}]_0}\right\}\\ &= \left\{x^\eta \tilde{h}(x^\gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathrm{SOS}([2\mathrm{d}]_0)\right\}.\end{aligned}$$

For $\gamma \in 2\mathbb{N}^{\mathrm{n}}$ the assertion follows from Proposition 4.3 and Corollary 4.12. Then $\tilde{\mathrm{K}} = \mathbb{R}_+$ and $\tilde{\mathrm{P}} = [2\mathrm{d}]_0$ and therefore

$$\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} = \left\{x^\eta \tilde{h}(x^\gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\mathbb{R}_+)_{[2\mathrm{d}]_0}\right\}$$

$$= \left\{ x^\eta \tilde{h}(x^\gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathrm{SOS}([2\mathrm{d}]_0) + y\,\mathrm{SOS}([2\mathrm{d}-2]_0) \right\}. \qquad \square$$

The characteristic feature of chains is that they reduce the multivariate case to the well-understood univariate case, that is $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} \cong \mathcal{P}(\mathbb{R})_{[2\mathrm{d}]_0}$ for $\gamma \notin 2\mathbb{N}^{\mathrm{n}}$ and $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} \cong \mathcal{P}(\mathbb{R}_+)_{[2\mathrm{d}]_0}$ for $\gamma \in 2\mathbb{N}^{\mathrm{n}}$.

Corollary 4.17 yields an analogue of SDSOS cones for chains: Recall that a polynomial f is SDSOS if

$$f \in \sum_{\mathrm{P} \in \mathcal{F}} \mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{P}}, \tag{4.3}$$

where $\mathcal{F}$ is a family of binomial square patterns. Due to Lemma 3.21 this can be enforced by PSD constraints. The $\mathbb{N}_{2\mathrm{d}}$-descriptions of $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ given in Corollary 4.17 allow to include other chains than binomial square patterns in $\mathcal{F}$ while maintaining that the constraint (4.3) can be enforced by PSD constraints of reasonable sizes. The matrices needed for these PSD constraints are not bigger than $(\mathrm{d}+1) \times (\mathrm{d}+1)$, which is indeed a reasonable dependency on d given that, currently, PSD solvers can handle dense $\mathrm{r} \times \mathrm{r}$ matrices for r well above 300 on every machine.

We proceed with the case when K is an axis-parallel box.

Corollary 4.18.
Let $\mathrm{CH}(\eta,\gamma,2\mathrm{d})$ *be a chain and* $\mathrm{K} = \mathrm{Box}(\mathbf{l},\mathbf{u})$ *with* $\mathbf{l} < \mathbf{u}$ *and* $\mathrm{d} \geq 1$. $\mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ *has a* $\mathbb{N}_{2\mathrm{d}}$*-description.* △

Proof. The proof follows from Proposition 4.3 together with either Theorem 4.13, Corollary 4.12 or Theorem 4.11.

The set $\tilde{\mathrm{P}}$ from Proposition 4.3 is $[2\mathrm{d}]_0$. In order to determine $\tilde{\mathrm{K}}$, let $\mathcal{I}$ be the set of subsets of $\mathbb{R}$ containing the sets of five types: $[\mathrm{a},\mathrm{b}]$, $(-\infty,\mathrm{a}]$, $[\mathrm{b},\infty)$, $(-\infty,\mathrm{a}] \cup [\mathrm{b},\infty)$ and $(-\infty,\infty)$, where $\mathrm{a},\mathrm{b} \in \mathbb{R}$ with $\mathrm{a} < \mathrm{b}$. Note that $\mathcal{I}$ is closed under multiplication, that is

$$\mathrm{I},\mathrm{J} \in \mathcal{I} \Rightarrow \mathrm{I} \cdot \mathrm{J} \in \mathcal{I}.$$

Furthermore, $\mathrm{cl}(\{\mathrm{x}_{\mathrm{i}}^{\gamma_{\mathrm{i}}} : \mathrm{x}_{\mathrm{i}} \in [\mathrm{l}_{\mathrm{i}},\mathrm{u}_{\mathrm{i}}] \backslash \{0\}\}) \in \mathcal{I}$. Hence, $\tilde{\mathrm{K}} = \mathrm{cl}(\{\mathbf{x}^\gamma : \mathbf{x} \in \mathrm{K} \text{ with } \mathrm{x}_{\mathrm{i}} \neq 0 \text{ for all } \mathrm{i} \in [\mathrm{n}]\}) \in \mathcal{I}$. In other words $\tilde{\mathrm{K}}$ corresponds to one of the five types of sets from above.

To conclude the proof, we show that there exist $\mathrm{r} \in \mathbb{N}$ and $g^{\mathrm{i}} \in \mathbb{R}[x], \mathrm{i} \in [\mathrm{r}]$ such that

$\mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} = \mathcal{S}(\{g^{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{r}]}, \mathbb{N}^{\mathrm{k}}_{[2\mathrm{d}]_0})$ by expressing $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ as

$$f = x^{\eta}\Big(s^0(x^{\gamma}) + \sum_{\mathrm{i}\in[\mathrm{r}]} s^{\mathrm{i}}(x^{\gamma})g^{\mathrm{i}}(x^{\gamma})\Big)$$

with $s^0 \in \mathrm{SOS}\,([2\mathrm{d}]_0)$ and $s^{\mathrm{i}} \in \mathrm{SOS}\,([2\mathrm{d}-2]_0)$ for all five types of $\tilde{\mathrm{K}}$.

If $\tilde{\mathrm{K}} = [\mathrm{a},\mathrm{b}]$, Theorem 4.13 asserts that $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ if and only if there exist $s^0 \in \mathrm{SOS}([2\mathrm{d}]_0)$ and $s^1 \in \mathrm{SOS}([2\mathrm{d}-2]_0)$ with

$$f = x^{\eta}\cdot\big(s^0(x^{\gamma}) + (x^{\gamma}-\mathrm{a})\cdot(\mathrm{b}-x^{\gamma})\cdot s^1(x^{\gamma})\big)\,.$$

If $\tilde{\mathrm{K}} = [\mathrm{b},\infty)$, then Corollary 4.12 states that

$$\mathcal{P}([\mathrm{b},\infty))_{[2\mathrm{d}]_0} = \mathrm{SOS}([2\mathrm{d}]_0) + (y-\mathrm{b})\cdot\mathrm{SOS}([2\mathrm{d}-2]_0).$$

Hence, $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ if and only if there exist SOS polynomials $s^0 \in \mathrm{SOS}([2\mathrm{d}]_0)$ and $s^1 \in \mathrm{SOS}([2\mathrm{d}-2]_0)$ with

$$f = x^{\eta}\cdot\big(s^0(x^{\gamma}) + (x^{\gamma}-\mathrm{b})\cdot s^1(x^{\gamma})\big)$$

Analogously, it follows for $\tilde{\mathrm{K}} = (-\infty,\mathrm{a}]$ that $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ if and only if there exist $s^0 \in \mathrm{SOS}([2\mathrm{d}]_0)$ and $s^1 \in \mathrm{SOS}([2\mathrm{d}-2]_0)$ with

$$f = x^{\eta}\cdot\big(s^0(x^{\gamma}) + (\mathrm{a}-x^{\gamma})\cdot s^1(x^{\gamma})\big)\,.$$

If $\tilde{\mathrm{K}} = (-\infty,\infty)$, Case (H1) of Theorem 4.11 asserts that $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ if and only if there exists $s^0 \in \mathrm{SOS}([2\mathrm{d}]_0)$ with

$$f = x^{\eta}\cdot s^0(x^{\gamma}).$$

Finally, if $\tilde{\mathrm{K}} = (-\infty,\mathrm{a}]\cup[\mathrm{b},\infty)$, Theorem 4.14 states that $f \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}$ if and only if there exist $s^0, s^1 \in \mathrm{SOS}([2\mathrm{d}]_0)$ and $s^1 \in \mathrm{SOS}([2\mathrm{d}-2]_0)$ such that

$$f = x^{\eta}\cdot\big(s^0(x^{\gamma}) + (\mathrm{a}-x^{\gamma})(\mathrm{b}-x^{\gamma})\cdot s^1(x^{\gamma})\big)\,. \qquad \square$$

With Corollary 4.18 we can extend the notion of SDSOS cones to boxes $\mathrm{K} = \mathrm{Box}(\mathbf{l},\mathbf{u})$, that is

$$f \in \sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_{\mathrm{P}} \tag{4.4}$$

where $\mathcal{F}$ is a family of chains (or just binomial square patterns). The $\mathbb{N}_{2\mathrm{d}}$-descriptions of $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ given in Corollary 4.18 allow to enforce (4.4) by PSD constraints. The maximal size of the matrices needed for the PSD constraints is $(\mathrm{d}+1)\times(\mathrm{d}+1)$.

Proposition 4.3 has the following consequences for the separation problem (SP) for chain $\mathrm{CH}(\gamma,\mathrm{d})$.

Remark 4.19 ([9]).
Let $\mathbf{l},\mathbf{u}\in\mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l}<\mathbf{u}$*,* $\mathrm{K}=\mathrm{Box}(\mathbf{l},\mathbf{u})$ *and* $p_{\mathbf{c}}(y) := \sum_{i=0}^{\mathrm{d}} \mathrm{c}_{\mathrm{i}\gamma} y^{\mathrm{i}}$*. Then*

$$\begin{array}{rlll} \text{maximize} & \delta - \langle\, \mathbf{c}, \mathbf{v}\,\rangle & & \\ \text{for} & \mathbf{c}\ \in \mathbb{R}^{\mathrm{CH}(\gamma,\mathrm{d})} & \text{and} & \delta\in\mathbb{R} \\ \text{subject to} & \mathrm{c}_\beta \in [-1,1] & \text{for all} & \beta\in\mathrm{CH}(\gamma,\mathrm{d}), \\ & p_{\mathbf{c}}(\mathrm{y}) \geq \delta & \text{for all} & \mathrm{y}\in[\underline{\mathbf{x}}^\gamma_{\mathrm{K}}, \overline{\mathbf{x}}^\gamma_{\mathrm{K}}]. \end{array} \tag{4.5}$$

Analogously to Corollary 4.18 it follows from Theorem 4.13 that the constraint $p_{\mathbf{c}}(y)-\delta\geq 0$ *on* $[\underline{\mathbf{x}}^\gamma_{\mathrm{K}}, \overline{\mathbf{x}}^\gamma_{\mathrm{K}}]$ *can be reformulated as a semidefinite constraint if* d *is even:*

$$\begin{array}{rl} \text{maximize} & \delta - \langle\, \mathbf{c}, \mathbf{v}\,\rangle \\ \text{for} & \mathbf{c}\ \in \mathbb{R}^{\mathrm{CH}(\eta,\gamma,\mathrm{d})},\ \delta\in\mathbb{R} \\ & s^0 \in \mathrm{SOS}([\mathrm{d}]_0)\ \text{ and }\ s^1\in\mathrm{SOS}([\mathrm{d}-2]_0) \\ \text{subject to} & \mathrm{c}_\beta\in[-1,1]\ \text{ for all }\ \beta\in\mathrm{CH}(\gamma,\mathrm{d}), \\ & p_{\mathbf{c}} - \delta = s^0 + s^1 g, \end{array} \tag{4.6}$$

where $g = (\overline{\mathbf{x}}^\gamma_{\mathrm{K}} - y)(y - \underline{\mathbf{x}}^\gamma_{\mathrm{K}})$*.* △

4.2 Discretized Chains

Another way to approach the separation problem for chains $\mathrm{CH}(\gamma,\mathrm{d})$ is by approximating the moment body $\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}$ by a polytope and solving a linear relaxation of (4.5). Proposition 4.20 shows how $\mathcal{M}(K)_{CH(\gamma,d)}$ can be approximated by polytopes to arbitrary precision. We will make use of the following notation. With each segment $[a,b] \subseteq \mathbb{R}$ we associate the $(d+1) \times (d+1)$ matrix

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

and the polytope

$$\Delta^{[a,b]} := \operatorname{conv}\{\Phi^{[a,b]}(\mathbf{w}^i) : i \in [d]_0\},$$

with $\mathbf{w}^i := \sum_{k=0}^{i} \mathbf{e}^k$ for $i \in [d]_0$. By a covering of an interval $[l,u]$ we understand a finite family of segments $\mathcal{I}$ satisfying $[l,u] = \bigcup_{I \in \mathcal{I}} I$. The fineness of $\mathcal{I}$ is defined as $\varrho(\mathcal{I}) := \max\{|b-a| : [a,b] \in \mathcal{I}\}$. Furthermore, let $\varepsilon > 0$ and $X \subseteq \mathbb{R}^A$ be a nonempty set. We call

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

the ε-neighbourhood of X and

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

the diameter of X.

Proposition 4.20 ([9]).
Let $d \in \mathbb{N}$ and $\mathcal{I}$ be a covering of segment $[l,u] \subseteq \mathbb{R}$. Then

$$\mathcal{M}([l,u])_{[d]_0} \subseteq \operatorname{conv}\left(\bigcup_{I \in \mathcal{I}} \Delta^I \right). \tag{4.7}$$

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

$$\operatorname{conv}\left(\bigcup_{I \in \mathcal{I}} \Delta^I \right) \subseteq N_\varepsilon(\mathcal{M}([l,u])_{[d]_0}). \tag{4.8}$$

△

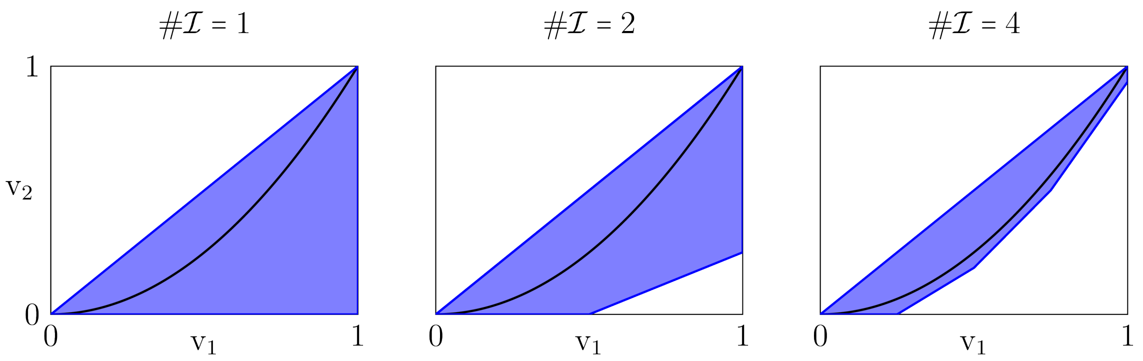

Figure 4.4: The plots show approximations $\bigcup_{I\in\mathcal{I}}\Delta^{I}$ of $\mathcal{M}([0,1])_{[2]}$ in blue and the curve $m([0,1])_{[2]}$ in black. **From left to right:** the coverings $\mathcal{I}$ contain an interval decomposition of $[0,1]$ into 1, 2 and 4 equidistant intervals.

Before giving a proof of Proposition 4.20, we establish the following.

Lemma 4.21 ([9]).
Let $d \in \mathbb{N}$, *let* $a, b \in \mathbb{R}$ *and* $a < b$. *Then one has*

$$\operatorname{diam}(\Delta^{[a,b]}) \leq |b-a| d^2 \eta_{a,b}^{d},$$

where

$$\eta_{a,b} := \max\{|a| + |b-a|, 1\}.$$

△

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

$$\operatorname{diam}(\Delta^{[a,b]}) = \max\{\|\Phi^{[a,b]}(\mathbf{w}^{j} - \mathbf{w}^{i})\|_1 : i, j \in [d]_0, i < j\}. \tag{4.9}$$

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

$$\mathbf{v}_k = \sum_{h=0}^{k} \binom{k}{h} a^{k-h}(b-a)^{h} w_h \tag{4.10}$$

for $k \in [d]_0$. The vector $\mathbf{w}$ belongs to $\{0,1\}^{[d]_0}$ and its component w_0 is 0. Since $w_0 = 0$, we obtain that $v_0 = 0$ and that the summand for $h = 0$ in the sum on the right hand side of (4.10) is zero. This yields

$$
\begin{aligned}
\|\mathbf{v}\|_1 &= |v_1| + \cdots + |v_d| \\
&\leq \sum_{k=1}^{d} \sum_{h=1}^{k} \binom{k}{h} |a|^{k-h} |b-a|^h \\
&= \sum_{k=1}^{d} \left((|a| + |b-a|)^k - |a|^k \right) \\
&\overset{(*)}{=} |b-a| \sum_{k=1}^{d} \sum_{h=0}^{k-1} (|a| + |b-a|)^h |a|^{k-1-h} \\
&\leq |b-a| \sum_{k=1}^{d} \sum_{h=0}^{k-1} \eta_{a,b}^{k-1} \\
&\leq |b-a| d^2 \eta_{a,b}^{d},
\end{aligned}
\tag{4.11}
$$

where the equation $(*)$ follows from the formula $t^k - q^k = (t-q) \sum_{h=0}^{k-1} q^h t^{k-1-h}$ for $t, q \in \mathbb{R}$. Applying the inequality (4.11) to (4.9) yields the assertion. □

Proof (Proposition 4.20). Consider an arbitrary segment $[a, b]$. We use the identity

$$
\Phi^{[a,b]} \, m(y)_{[d]_0} = m((1-y)a + yb)_{[d]_0}, \tag{4.12}
$$

which holds for every $y \in \mathbb{R}$ and can be derived using the binomial expansion for the components of the right-hand side of (4.12), that is for $k \in [d]_0$

$$
\begin{aligned}
m((1-y)a + yb)_k &= (a + (b-a)y)^k \\
&= \sum_{j=0}^{k} \binom{k}{j} a^{k-j} (b-a)^j y^j \\
&= [\![\Phi^{[a,b]} \, m(y)_{[d]_0}]\!]_k.
\end{aligned}
$$

The polytope $\operatorname{conv}(\{\mathbf{w}^0, \ldots, \mathbf{w}^d\})$ is a d-dimensional simplex,[6] which can be described as the set of all $\mathbf{v} \in \mathbb{R}^{[d]_0}$ satisfying $1 = v_0 \geq v_1 \geq \ldots \geq v_d \geq 0$. Hence, $m(y)_{[d]_0}$ belongs to $\operatorname{conv}(\{\mathbf{w}^0, \ldots, \mathbf{w}^d\})$ for every $y \in [0, 1]$. Consequently, $\Phi^{[a,b]} \, m(y)_{[d]_0}$ belongs

[6] Recall that $\mathbf{w}^i = \sum_{k=0}^{i} \mathbf{e}^k$.

to $\operatorname{conv}(\{\Phi^{[a,b]}\mathbf{w}^0, \ldots, \Phi^{[a,b]}\mathbf{w}^d\}) = \Delta^{[a,b]}$. Taking into account (4.12), we see that

$$\mathrm{m}(y)_{[d]_0} \in \Delta^{[a,b]}$$

holds for every $y \in [a, b]$. The latter immediately implies (4.7).

We now derive the second part of the assertion. In view of Lemma 4.21, if $[a, b] \in \mathcal{I}$, then every point of $\Delta^{[a,b]}$ has l_1-distance at most $|b - a| d^2 \eta_{a,b}^d$ to the point $\mathrm{m}(y)_{[d]_0}$, which belongs to $\mathcal{M}([l, u])_{[d]_0}$ and to $\Delta^{[a,b]}$. This yields

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

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

$$\bigcup_{[a,b] \in \mathcal{I}} \Delta^{[a,b]} \subseteq \mathrm{N}_\varepsilon(\mathcal{M}([l, u])_{[d]_0}) \tag{4.13}$$

for every $\varepsilon \geq \frac{\varrho(\mathcal{I})}{C^\star}$ with $C^\star = (d^2 \eta_{l,u}^d)^{-1}$. 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 (4.8) holds if $\varepsilon > 0$ satisfies $\rho(\mathcal{I}) \geq C^\star \varepsilon$. □

Proposition 4.20 allows to solve the separation problem for $\mathcal{M}(K)_{CH(\gamma,[d]_0)}$ approximately using linear programming.

Corollary 4.22 ([9]).
Let $K = \operatorname{Box}(\mathbf{l}, \mathbf{u})$ *with* $\mathbf{l} < \mathbf{u}$, $d \geq 1$ *and* $CH(\gamma, 2d)$. *Then there exists a constant* $C^\star > 0$ *that depends only on* $\underline{\mathbf{x}}_K^\gamma, \overline{\mathbf{x}}_K^\gamma$ *and* d *such that the following holds: If a vector* $\mathbf{v} \in \mathbb{R}^{CH(\gamma,d)}$ *does not belong to* $\mathcal{M}(K)_{CH(\gamma,d)}$ *and* $\mathcal{I}$ *is a covering of* $[\underline{\mathbf{x}}_K^\gamma, \overline{\mathbf{x}}_K^\gamma]$ *with*

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

then the optimal value of the linear program

$$\begin{aligned}
\text{maximize} \quad & \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
\text{for} \quad & \mathbf{c} \in \mathbb{R}^{CH(\eta,\gamma,d)} \text{ and } \delta \in \mathbb{R} \\
\text{subject to} \quad & c_\beta \in [-1, 1] \qquad \text{for all } \beta \in CH(\eta, \gamma, d), \\
& \sum_{i=0}^{d} c_{\eta+i\gamma} [\![\Phi^I \mathbf{w}^j]\!]_i \geq \delta \quad \text{for all } I \in \mathcal{I} \text{ and } j \in [d]_0
\end{aligned} \tag{4.14}$$

is positive. △

Proof. By Proposition 4.10,

$$\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})} = \mathcal{M}([\underline{\mathbf{x}}_{\mathrm{K}}^{\gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\gamma}])_{[\mathrm{d}]_0}.$$

We choose $\mathrm{C}^\star > 0$ as in Proposition 4.20 for $[\mathrm{l}, \mathrm{u}] = [\underline{\mathbf{x}}_{\mathrm{K}}^{\gamma}, \overline{\mathbf{x}}_{\mathrm{K}}^{\gamma}]$ and fix $\varepsilon := \varrho(\mathcal{I})/\mathrm{C}^\star$. Since

$$\mathrm{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}, \mathbf{v}) > \varepsilon,$$

the vector $\mathbf{v}$ does not belong to $\mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})})$. Hence, in view of Proposition 4.20,

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

By separation theorems, there exists a vector $\mathbf{c} \in \mathbb{R}^{\mathrm{CH}(\gamma,\mathrm{d})}$ with $\|\mathbf{c}\|_\infty \le 1$ and $\delta \in \mathbb{R}$ such that $\langle \mathbf{c}, \mathbf{v}\rangle < \delta$ and $\langle \mathbf{c}, \mathbf{u}\rangle \ge \delta$ for all $\mathbf{u} \in \mathrm{conv}\left(\bigcup_{\mathrm{I}\in\mathcal{I}} \Delta^{\mathrm{I}}\right)$. Hence, $\mathbf{c}$ and δ are feasible for (4.14) and their corresponding objective value is positive. □

4.3 Shifted Chain Patterns

Definition 4.23 (Shifted Chain Pattern).
Let $\eta \in 2\mathbb{N}^{\mathrm{n}}$, $\gamma \in \mathbb{Z}^{\mathrm{n}}$ *with* $\gamma \neq \mathbf{0}$ *and* $\mathrm{d} \in \mathbb{N}\backslash\{0\}$ *satisfy* $\eta + \mathrm{d}\gamma \ge \mathbf{0}$. *Furthermore, let* $\xi \in \mathbb{N}^{\mathrm{n}}$ *with* $\mathrm{supp}(\gamma) \cap \mathrm{supp}(\xi) = \emptyset$ *and* $\mathrm{supp}(\eta) \cap \mathrm{supp}(\xi) = \emptyset$. *We call*

$$\xi + \mathrm{CH}(\eta, \gamma, \mathrm{d})$$

a shifted chain. △

Using Proposition 2.13 we can represent the moment body $\mathcal{M}(\mathrm{K})_{\xi+\mathrm{CH}(\gamma,\mathrm{d})}$ for $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ as the convex hull of $\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}$ and $\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}$. Figure 4.5 illustrates this with a small example.

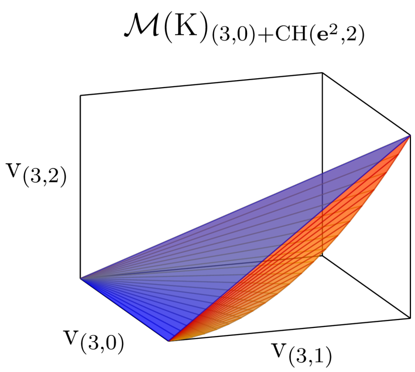

Figure 4.5: The moment body $\mathcal{M}(\mathrm{K})_{(3,0)+\mathrm{CH}(\mathbf{e}^2,2)}$ is shown for $\mathrm{K} = [0,1]^2$, that is the convex hull of $\mathbf{0}$ and $\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\mathbf{e}^2,2)}$. The set $\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\mathbf{e}^2,2)}$ is highlighted in orange. Note that $(3,0) + \mathrm{CH}(\mathbf{e}^2,2)) = \{(3,0),(3,1),(3,2)\}$.

We formulate analogous results for Remark 4.19 and Corollary 4.22 for shifted chains.

Corollary 4.24 ([9]).
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l} < \mathbf{u}$, $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ *and* $g = (\overline{\mathbf{x}}_{\mathrm{K}}^{\gamma} - y)(y - \underline{\mathbf{x}}_{\mathrm{K}}^{\gamma})$. *Then problem* (SP) *for the pattern* $\xi + \mathrm{CH}(\gamma, 2\mathrm{d})$ *can be formulated as*

$$\begin{array}{rl} \text{maximize} & \delta - \langle\, \mathbf{c}, \mathbf{v} \,\rangle \\ \text{for} & \mathbf{c} \ \in \mathbb{R}^{\xi+\mathrm{CH}(\gamma,\mathrm{d})},\ \delta \in \mathbb{R}, \\ & s^0, \tilde{s}^0 \in \mathrm{SOS}(\mathbb{N}_{2\mathrm{d}}) \ \text{ and } \ s^1, \tilde{s}^1 \in \mathrm{SOS}(\mathbb{N}_{2\mathrm{d}-2}) \\ \text{subject to} & \mathrm{c}_\beta \ \in [-1,1] \ \text{ for all } \ \beta \in \mathrm{CH}(\gamma, \mathrm{d}), \\ & \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, p_{\mathbf{c}} - \delta = s^0 + s^1 g, \\ & \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, p_{\mathbf{c}} - \delta = \tilde{s}^0 + \tilde{s}^1 g, \end{array}$$

where $p_{\mathbf{c}}(y) := \sum_{i=0}^{2\mathrm{d}} \mathrm{c}_{\mathrm{i}\gamma} y^{\mathrm{i}}$. △

Corollary 4.25 ([9]).
Let $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$ *with* $\mathbf{l} < \mathbf{u}$ *and* $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$. *Then there exists a constant* $\mathrm{C} > 0$ *that depends only on* $\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}, \overline{\mathbf{x}}_{\mathrm{K}}^{\gamma}$ *and* d *such that the following holds: If a vector* $\mathbf{v} \in \mathbb{R}^{\xi+\mathrm{CH}(\gamma,\mathrm{d})}$ *does not belong to* $\mathcal{M}(\mathrm{K})_{\xi+\mathrm{CH}(\gamma,\mathrm{d})}$ *and* $\mathcal{I}$ *is a covering of* $[\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}, \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}]$ *with*

$$\varrho(\mathcal{I}) < \mathrm{C}\,\mathrm{dist}(\mathcal{M}(\mathrm{K})_{\xi+\mathrm{CH}(\gamma,\mathrm{d})}, \mathbf{v}),$$

then the optimal value of the linear program

$$
\begin{aligned}
\text{maximize} \quad & \delta - \langle \mathbf{c}, \mathbf{v} \rangle \\
\text{for} \quad & \mathbf{c} \in \mathbb{R}^{\xi+\mathrm{CH}(\gamma,\mathrm{d})} \ \text{ and } \ \delta \in \mathbb{R} \\
\text{subject to} \quad & \mathrm{c}_\beta \in [-1,1] && \text{for all } \ \beta \in \xi + \mathrm{CH}(\gamma,\mathrm{d}), \\
& \sum_{\mathrm{i}\in[\mathrm{d}]_0} \mathrm{c}_{\xi+\mathrm{i}\gamma} [\![\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \Phi^{\mathrm{I}} \mathbf{u}^{\mathrm{j}}]\!]_{\mathrm{i}} \geq \delta && \text{for all } \ \mathrm{I} \in \mathcal{I} \ \text{ and } \ \mathrm{j} \in [\mathrm{d}]_0, \\
& \sum_{\mathrm{i}\in[d]_0} \mathrm{c}_{\eta+\mathrm{i}\gamma} [\![\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \Phi^{\mathrm{I}} \mathbf{u}^{\mathrm{j}}]\!]_{\mathrm{i}} \geq \delta && \text{for all } \ \mathrm{I} \in \mathcal{I} \ \text{ and } \ \mathrm{j} \in [\mathrm{d}]_0
\end{aligned}
\tag{4.15}
$$

is positive. △

Proof. Let $\mathrm{C}^\star$ be the constant from Proposition 4.20 for $[\mathrm{a},\mathrm{b}] = [\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}, \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}]$ and $\kappa := \max\{|\,\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\,|, |\,\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\,|\}$. Using Proposition 2.13 we write

$$\mathcal{M}(\mathrm{K})_{\xi+\mathrm{CH}(\gamma,\mathrm{d})} = \mathrm{conv}(\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})} \cup \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}).$$

We claim that

$$
\begin{aligned}
&\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}) \cup \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}) \subseteq \\
&\mathrm{N}_{\kappa\varepsilon}(\mathrm{conv}(\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})} \cup \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})})).
\end{aligned}
$$

For the proof of the claim let $\mathbf{v} \in \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})})$. Then there exists $\mathbf{w} \in \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}$ with $\|\frac{\mathbf{v}}{\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}} - \mathbf{w}\|_1 \leq \varepsilon$. Hence, $\|\mathbf{v} - \underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathbf{w}\|_1 \leq |\,\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\,|\varepsilon$ and therefore

$$\mathbf{v} \in \mathrm{N}_{|\,\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\,|\varepsilon}(\underline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})}).$$

Analogously, if $\mathbf{v} \in \overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})})$, then $\mathbf{v} \in \mathrm{N}_{|\,\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\,|\varepsilon}(\overline{\mathbf{x}}_{\mathrm{K}}^{\xi}\, \mathcal{M}(\mathrm{K})_{\mathrm{CH}(\gamma,\mathrm{d})})$, which concludes the proof of the claim.

The assertion follows with $\mathrm{C} = \frac{\mathrm{C}^\star}{\kappa}$ using the same arguments as in the proof of Corollary 4.22. □

4.4 Multivariate Quadratic Patterns and Bivariate Quartic Patterns

The next two patterns are further examples that show how to construct new pattern types from existing Positivstellensätze.

Definition 4.26 (Multivariate Quadratic Patterns).
Let $\eta \in 2\mathbb{N}^n$ and $\Gamma = (\gamma^1, \ldots, \gamma^k) \in \mathbb{Z}^{n\times k}$ with $k \in [n]$ be a matrix, whose columns γ^i are nonzero vectors with pairwise disjoint supports that satisfy $\eta + 2\gamma^i \in \mathbb{N}^n$. We call

$$\mathrm{MQ}(\eta, \Gamma) := \left\{\eta + \gamma^1\omega_1 + \cdots + \gamma^k\omega_k : \omega \in \mathbb{N}_2^k\right\},$$

a k-variate quartic pattern (MQ) with shift η.[7] △

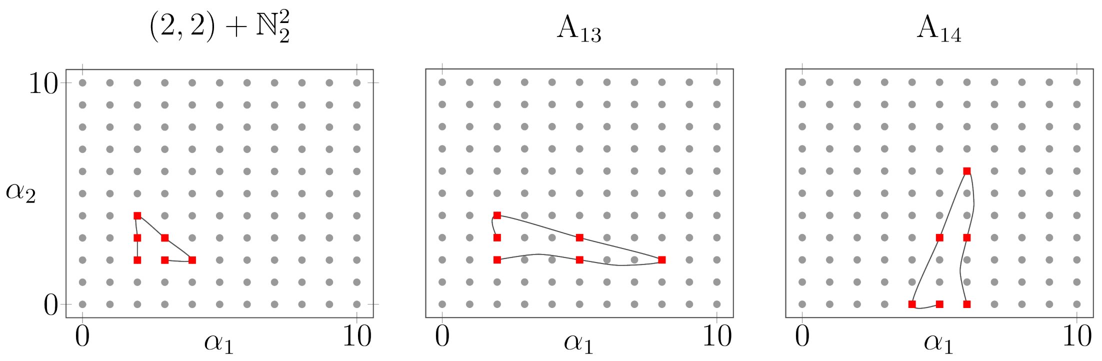

Figure 4.6: Visualization of possible quartic patterns as described in Plot Set Up 3.2. **Left:** $\mathrm{MQ}((2,2), \mathbf{e}^1, \mathbf{e}^2)$. **Middle:** $\mathrm{MQ}((2,2), 3\cdot\mathbf{e}^1, \mathbf{e}^2)$. **Right:** $\mathrm{MQ}((6,0), -\mathbf{e}^1, 3\cdot\mathbf{e}^2)$.

Definition 4.27 (Bivariate Quartic Patterns).
Let $\eta \in 2\mathbb{N}^n$ and $\gamma^1, \gamma^2 \in \mathbb{Z}^n$ be nonzero vectors with pairwise disjoint supports that satisfy $\eta + 4\gamma^i \in \mathbb{N}^n$ for $i \in [2]$. We call

$$\mathrm{BQ}(\eta, \gamma^1, \gamma^2) := \left\{\eta + \gamma^1\omega_1 + \gamma^2\omega_2 : \omega \in \mathbb{N}_4^2\right\},$$

a bivariate quartic pattern (BQ) with shift η.[7] △

[7] Whenever we write $\mathrm{MQ}(\eta, \Gamma)$ or $\mathrm{BQ}(\eta, \gamma^1, \gamma^2)$ we imply that the parameters satisfy the assumptions of Definition 4.26 respectively Definition 4.27.

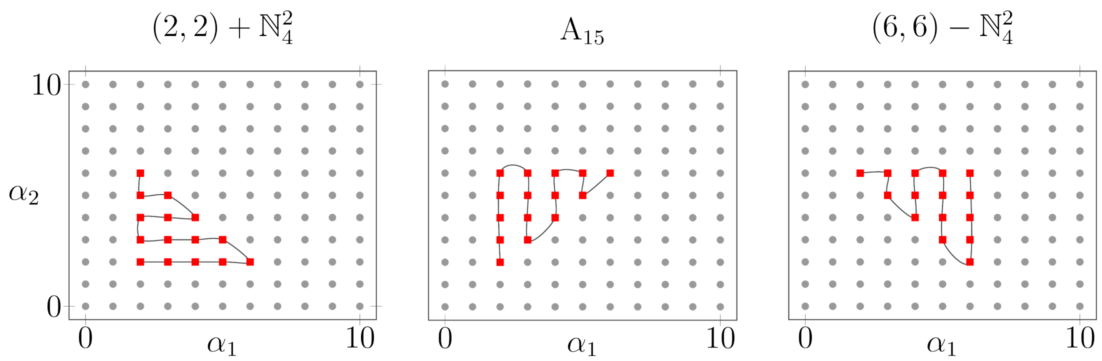

Figure 4.7: Visualization of possible bivariate quartic patterns as described in Plot Set Up 3.2. **Left:** $\mathrm{BQ}((2,2),\mathbf{e}^1,\mathbf{e}^2)$. **Middle:** $\mathrm{BQ}((2,6),\mathbf{e}^1,-\mathbf{e}^2)$. **Right:** $\mathrm{BQ}((6,6),-\mathbf{e}^1,-\mathbf{e}^2)$.

Corollary 4.28.
If the generators γ^{i} of $\mathrm{MQ}(\eta,\Gamma)$ with $\Gamma = (\gamma^1,\ldots,\gamma^{\mathrm{k}})$ satisfy $\gamma^{\mathrm{i}} \notin 2\mathbb{Z}^{\mathrm{n}}$, then there exists an $\mathbb{N}_2^{\mathrm{k}}$-representation of $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{MQ}(\eta,\Gamma)}$.

If the generators $\mathrm{BQ}(\eta,\gamma^1,\gamma^2)$ satisfy $\gamma^1,\gamma^2 \notin 2\mathbb{Z}^{\mathrm{n}}$ then there exists a $\mathbb{N}_4^2$-representation of $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{MQ}(\eta,\Gamma)}$.

Proof. Both cases follow from Proposition 4.3 by determining the respective $\tilde{\mathrm{K}}$ and $\tilde{\mathrm{P}}$. The first follows with $\tilde{\mathrm{K}} = \mathbb{R}^{\mathrm{k}}$ and $\tilde{\mathrm{P}} = \mathbb{N}_2^{\mathrm{k}}$ from the case (H2) of Theorem 4.11:

$$\begin{aligned}\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{MQ}(\eta,\Gamma)} &= \left\{x^\eta\tilde{h}(\mathrm{m}(x)_\Gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\mathbb{R}^{\mathrm{k}})_{\mathbb{N}_2^{\mathrm{k}}}\right\}\\ &= \left\{x^\eta\tilde{h}(\mathrm{m}(x)_\Gamma) \in \mathbb{R}[x] : \tilde{h} \in \mathrm{SOS}(\mathbb{N}_2^{\mathrm{k}})\right\}.\end{aligned}$$

Analogously, the second follows with $\tilde{\mathrm{K}} = \mathbb{R}^2$ and $\tilde{\mathrm{P}} = \mathbb{N}_4^2$ from the case (H3) of Theorem 4.11:

$$\begin{aligned}\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathrm{BQ}(\eta,\gamma^1,\gamma^2)} &= \left\{x^\eta\tilde{h}\left(\mathrm{m}(x)_{\{\gamma^1,\gamma^2\}}\right) \in \mathbb{R}[x] : \tilde{h} \in \mathcal{P}(\mathbb{R}^2)_{\mathbb{N}_4^2}\right\}\\ &= \left\{x^\eta\tilde{h}\left(\mathrm{m}(x)_{\{\gamma^1,\gamma^2\}}\right) \in \mathbb{R}[x] : \tilde{h} \in \mathrm{SOS}(\mathbb{N}_4^2)\right\}.\end{aligned}$$ □

Chapter 5

Algorithms

The first part of this chapter is devoted to the development of the algorithms used for solving a pattern relaxation of (POP). In the last section we present an idea of a graph based approach for the selection of patterns and discuss it by considering an example.

In order to solve the pattern relaxations two different approaches were chosen: the first is a hybrid cutting-plane algorithm for (P-RLX') and the second uses a conic programming solver to compute (D-RLX). The hybrid cutting-plane algorithm starts with a relaxation of (P-RLX') and refines this relaxation by iteratively including inequalities that stem from the separation problems of the various pattern types. Consequently, a convergence proof of the procedure is provided. As the separation problems in each iteration are independent of each other, solving them can be parallelized. However, this possibility was not explored within this thesis. Furthermore, if the initial relaxation is chosen to be polyhedral, the subsequent relaxations constructed in each iteration are LPs while separation problems may be SDPs or formulations utilizing power cones. This yields the possibility to warm start each iteration – something that is usually not feasible for other solution methods. Both methods have been implemented in `Matlab` [41] and use as subsolver `MOSEK`. The respective programs can be found on the CD located in the back cover of this thesis.

5.1 Cutting-Plane Algorithm

Definition 5.1 (ε-feasible and ε-optimal).
Consider an optimization problem

$$\min\{f(\mathbf{x}) : \mathbf{x} \in \mathrm{K}\} \tag{5.1}$$

with the optimal value $\mathrm{f}^\star$. *Let* $\varepsilon > 0$. *We say that* $\mathbf{x} \in \mathbb{R}^{\mathrm{n}}$ *is* ε-feasible *for* (5.1) *if* $\mathbf{x} \in \mathrm{N}_\varepsilon(\mathrm{K})$ *and* ε-optimal *for* (5.1) *if* $\mathrm{f}^\star - \varepsilon \leq f(\mathbf{x}) \leq \mathrm{f}^\star + \varepsilon$. $\triangle$

In Algorithm 5.1 we formulate a hybrid cutting-plane algorithm for solving (P-RLX'), when $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ with $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$, $\mathbf{l} < \mathbf{u}$. The underlying idea is to divide $\mathcal{F}$ into statically used patterns $\mathcal{F}^{'}$ and dynamically used patterns $\mathcal{F}^{''}$. The basic constraints $\mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}}, \mathrm{P} \in \mathcal{F}^{'}$ are always considered in the master problem (MP), while cuts are generated by solving the slave (or separation) problems (SP) for more involved patterns $\mathrm{P} \in \mathcal{F}^{''}$. This way $\mathcal{F}^{'}$ and $\mathcal{F}^{''}$ allow to control the size and type of the master problem.

In particular, choosing $\mathcal{F}^{'}$ to contain only multilinear patterns makes (MP) a linear program, while the slave problems may be SDPs. This allows to solve (MP) with the dual simplex algorithm, warmstarting after the inclusion of linear cuts generated from $\mathrm{P} \in \mathcal{F}^{''}$. Similarly, in a branch-and-bound framework cuts corresponding to patterns $\mathrm{P} \in \mathcal{F}^{''}$ could be used to augment the pattern relaxation (MP) until the bounds are tight enough for further branching.

Algorithm 5.1: Hybrid Cutting-Plane Algorithm

Input: • finite family of patterns $\mathcal{F} = \mathcal{F}^{'} \cup \mathcal{F}^{''}$, where $\mathcal{F}^{'}$ contains statically and $\mathcal{F}^{''}$ contains dynamically used patterns • coefficient vector $\mathbf{f} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}}$ • tolerance $\varepsilon > 0$ • $\mathrm{K} = \mathrm{Box}(\mathbf{l}, \mathbf{u})$ with $\mathbf{l}, \mathbf{u} \in \mathbb{R}^{\mathrm{n}}$, $\mathbf{l} < \mathbf{u}$
Output: $\mathcal{O}(\varepsilon)$-optimal and $\mathcal{O}(\varepsilon)$-feasible solution $\mathbf{v}^{\star}(\varepsilon)$ of (P-RLX').

(0) Set $\mathrm{i} = 0$.
(1) Solve the following problem

$$\begin{array}{lll} \text{maximize} & \langle\, \mathbf{f}, \mathbf{v} \,\rangle & \\ \text{for} & \mathbf{v} \ \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} & \\ \text{subject to} & \mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}} & \text{for all } \mathrm{P} \in \mathcal{F}^{'}, \\ & \mathrm{v}_{\alpha} \in [\underline{\mathbf{x}}_{\mathrm{K}}^{\alpha}, \overline{\mathbf{x}}_{\mathrm{K}}^{\alpha}] & \text{for all } \alpha \in \mathrm{A}_{\mathcal{F}}, \\ & \mathbf{v} \ \text{ satisfies } \mathrm{Ineq}^{\mathrm{k}} & \text{for all } \mathrm{k} \in [\mathrm{i}] \end{array} \tag{MP}$$

and save the minimizer as $\mathbf{v}^{(\mathrm{i}+1)}$. Set $\mathrm{i} \leftarrow \mathrm{i} + 1$. Initialize $\mathrm{Ineq}^{\mathrm{i}} = \emptyset$.
(2) For each pattern $\mathrm{P} \in \mathcal{F}^{''}$ determine the corresponding distance $\delta := \mathrm{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \mathbf{v}_{\mathrm{P}}^{(\mathrm{i})})$ by solving the respective separation problem (SP). If $\delta > \varepsilon$, add $\langle \mathbf{c}^{\mathrm{P}}, \mathbf{v}_{\mathrm{P}} \rangle \geq \delta$ to the set of inequalities $\mathrm{Ineq}^{\mathrm{i}}$.
(3) If the set $\mathrm{Ineq}^{\mathrm{i}}$ is empty, return $\mathbf{v}^{\star}(\varepsilon) := \mathbf{v}^{(\mathrm{i})}$, else go to step (1).

We are now proving finite termination and a bound on the result of Algorithm 5.1.

Theorem 5.2 ([9]).
For every given $\varepsilon > 0$, Algorithm 5.1 terminates after a finite number of iterations. The output satisfies

$$\begin{aligned} \mathbf{v}_{\mathrm{P}}^{\star} &\in \mathcal{M}(\mathrm{K})_{\mathrm{P}} && \text{for all } \mathrm{P} \in \mathcal{F}' \\ \mathbf{v}_{\mathrm{P}}^{\star}(\varepsilon) &\in \mathrm{N}_{\varepsilon}(\mathcal{M}(\mathrm{K}))_{\mathrm{P}} && \text{for all } \mathrm{P} \in \mathcal{F}''. \end{aligned} \qquad \triangle$$

Proof. Assume that Algorithm 5.1 does not terminate after a finite number of iterations. Then it produces an infinite sequence $\{\mathbf{v}^{(\mathrm{i})}\}_{\mathrm{i}\in\mathbb{N}}$ such that for all i there exists a $\mathrm{P}^{(\mathrm{i})} \in \mathcal{F}''$ with

$$\mathbf{v}_{\mathrm{P}^{(\mathrm{i})}} \notin \mathrm{N}_{\varepsilon}(\mathcal{M}(\mathrm{K})_{\mathrm{P}^{(\mathrm{i})}}).$$

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

$$\operatorname{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})}) > \varepsilon.$$

for all $\mathrm{i} \in \mathbb{N}$. Hence, for i large enough, $\|\tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})} - \tilde{\mathbf{v}}_{\mathrm{P}}\|_1 < \operatorname{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})})$ holds. Application of Proposition 2.12 to the minimizers $\mathbf{c}^{(\mathrm{i})}, \delta^{(\mathrm{i})}$ of (SP) for $\mathbf{v} = \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})}$ and the Hölder inequality yield

$$\begin{aligned} \langle \mathbf{c}^{(\mathrm{i})}, \tilde{\mathbf{v}}_{\mathrm{P}} \rangle &= \langle \mathbf{c}^{(\mathrm{i})}, \tilde{\mathbf{v}}_{\mathrm{P}} - \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})} \rangle + \langle \mathbf{c}^{(\mathrm{i})}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})} \rangle \\ &= \langle \mathbf{c}^{(\mathrm{i})}, \tilde{\mathbf{v}}_{\mathrm{P}} - \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})} \rangle + \delta^{(\mathrm{i})} - \operatorname{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})}) \\ &< \operatorname{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})}) + \delta^{(\mathrm{i})} - \operatorname{dist}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}, \tilde{\mathbf{v}}_{\mathrm{P}}^{(\mathrm{i})}) = \delta^{(\mathrm{i})}. \end{aligned}$$

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

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

Theorem 5.3.
There exists a constant $\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})$ depending on K and $\mathrm{A}_{\mathcal{F}}$ such that for every $\varepsilon > 0$ the distance between the feasible set of (P-RLX') *and the output $\mathbf{v}^{\star}(\varepsilon)$ of Algorithm 5.1*

is at most $\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon$. *Furthermore, the optimal value* $\mathrm{f}^\star$ *of* (P-RLX') *satisfies*

$$\langle \mathbf{f}, \mathbf{v}^\star(\varepsilon)\rangle \leq \mathrm{f}^\star \leq \|\mathbf{f}\|_\infty \mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon + \langle \mathbf{f}, \mathbf{v}^\star(\varepsilon)\rangle.$$

△

For the proof we use the following lemma, which is an adaptation of [56, Lem. 1.8.9].

Lemma 5.4 ([9]).
Let $\mathrm{X}, \mathrm{X}^1, \ldots, \mathrm{X}^\mathrm{r}$ *be nonempty compact convex subsets of* $\mathbb{R}^\mathrm{A}$ *such that the intersection* $\mathrm{Y} := \mathrm{X}^1 \cap \ldots \cap \mathrm{X}^\mathrm{r}$ *contains an* l_1*-ball of radius* $\rho > 0$ *as a subset and the inclusion* $\mathrm{X}^\mathrm{i} \subseteq \mathrm{X}$ *holds for every* $\mathrm{i} \in [\mathrm{r}]$. *Let* $\varepsilon > 0$ *and let* $\mathbf{x}$ *be a point of* X *satisfying* $\mathrm{dist}(\mathrm{X}^\mathrm{i}, \mathbf{x}) \leq \varepsilon$ *for every* $\mathrm{i} \in [\mathrm{r}]$. *Then*

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

△

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

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

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

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

Since $\|\mathbf{x} - \mathbf{x}^\mathrm{i}\|_1 \leq \varepsilon$, we immediately get $\mathbf{p}^\mathrm{i} \in \mathrm{B}$. The proof is concluded by estimating the distance between $\mathbf{x}$ and $\mathbf{y}$

$$\begin{aligned}\mathrm{dist}(\mathrm{Y}, \mathbf{x}) &\leq \|\mathbf{y} - \mathbf{x}\|_1 \\ &= \frac{\varepsilon}{\rho+\varepsilon}\|\mathbf{x}\|_1.\end{aligned}$$

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 $\|\mathbf{x}\|_1 \leq \mathrm{diam}(\mathrm{X})$. Thus, we arrive at the desired estimate of $\mathrm{dist}(\mathrm{Y}, \mathbf{x})$. $\square$

Proof (of Theorem 5.3). If some pattern $\mathrm{P} \in \mathcal{F}$ contains $\mathbf{0}$, then $\mathbf{0} \in \mathrm{A}_{\mathcal{F}}$ and the constraint $\mathrm{v}_\alpha \in [\underline{\mathbf{x}}_{\mathrm{K}}^{\alpha}, \overline{\mathbf{x}}_{\mathrm{K}}^{\alpha}]$ with $\alpha = 0$ occurring in Ineq^0 can be formulated as the equality $\mathrm{v}_{\mathbf{0}} = 1$. This shows that all solutions generated during the iteration satisfy the constraint $\mathrm{v}_{\mathbf{0}} = 1$. Using this observation and Theorem 5.2, which yields $\mathbf{v}_{\mathrm{P}}^{\star}(\varepsilon) \in \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{\mathrm{P}})$, we obtain $\mathbf{v}_{\mathrm{P}\setminus\{\mathbf{0}\}}^{\star}(\varepsilon) \in \mathrm{N}_\varepsilon(\mathcal{M}(\mathrm{K})_{P\setminus\{\mathbf{0}\}})$. We can therefore remove $\mathbf{0}$ from all patterns and assume that $\mathbf{0} \notin \mathrm{P}$ holds for every $\mathrm{P} \in \mathcal{F}$. The feasible set of (P-RLX') is the intersection of

$$\mathrm{F}^{\mathrm{P}} := \{\mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} : \mathbf{v}_{\mathrm{P}} \in \mathcal{M}(\mathrm{K})_{\mathrm{P}}\} \cap \mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}})$$

for all $\mathrm{P} \in \mathcal{F}$. We show that $\mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}$ is full-dimensional by assuming the contrary. Then $\mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}$ is contained in a linear subspace $\{\mathbf{v} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} : \langle \mathbf{c}, \mathbf{v} \rangle = 0\}$ given by $\mathbf{c} \in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}}$ with $\mathbf{c} \neq \mathbf{0}$. Hence, $\langle \mathbf{c}, \mathrm{m}(\mathbf{x})_{\mathrm{A}_{\mathcal{F}}} \rangle$ is a polynomial in $\mathbf{x}$ vanishing on a n-dimensional set K. This implies $\mathbf{c} = \mathbf{0}$, which is a contradiction. Hence, $\mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}$ is full-dimensional and therefore contains a ball with radius $\mathrm{R}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})$ depending only on K and $\mathrm{A}_{\mathcal{F}}$. The feasible set $\bigcap_{\mathrm{P} \in \mathcal{F}} \mathrm{F}^{\mathrm{P}}$ of (P-RLX') contains $\mathcal{M}(\mathrm{K})_{\mathrm{A}_{\mathcal{F}}}$ and therefore the aforementioned ball. The diameter $\mathrm{D}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})$ of $\mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}})$ depends only on K and $\mathrm{A}_{\mathcal{F}}$. From $\mathbf{v}^{\star}(\varepsilon) \in \bigcap_{\mathrm{P} \in \mathcal{F}'} \mathrm{F}^{\mathrm{P}} \cap \bigcap_{\mathrm{P} \in \mathcal{F}''} \mathrm{N}_\varepsilon(\mathrm{F}^{\mathrm{P}}) \cap \mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}})$ and Lemma 5.4 it follows that

$$\mathrm{dist}\Big(\bigcap_{\mathrm{P} \in \mathcal{F}} (\mathrm{F}^{\mathrm{P}} \cap \mathrm{Box}(\underline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}}, \overline{\mathbf{x}}_{\mathrm{K}}^{\mathrm{A}_{\mathcal{F}}})), \mathbf{v}^{\star}(\varepsilon)\Big) \leq \underbrace{\frac{\mathrm{D}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})}{\mathrm{R}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})}}_{:=\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})} \varepsilon.$$

At last, let $\mathbf{v} \in \bigcap_{\mathrm{P} \in \mathcal{F}} \mathrm{F}^{\mathrm{P}}$ satisfy $\|\mathbf{v} - \mathbf{v}^{\star}(\varepsilon)\|_1 \leq \mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon$. The upper bound on $\mathrm{f}^{\star}$ in terms of $\mathbf{v}^{\star}(\varepsilon)$ follows from

$$\begin{aligned} \mathrm{f}^{\star} &\leq \langle \mathbf{f}, \mathbf{v} \rangle \\ &= \langle \mathbf{f}, \mathbf{v} - \mathbf{v}^{\star}(\varepsilon) \rangle + \langle \mathbf{f}, \mathbf{v}^{\star}(\varepsilon) \rangle \\ &\leq \|\mathbf{f}\|_\infty \|\mathbf{v}^{\star} - \mathbf{v}^{\star}(\varepsilon)\|_1 + \langle \mathbf{f}, \mathbf{v}^{\star}(\varepsilon) \rangle \\ &= \|\mathbf{f}\|_\infty \mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon + \langle \mathbf{f}, \mathbf{v}^{\star}(\varepsilon) \rangle. \end{aligned}$$

$\square$

If the separation problem for some patterns P is too hard, it might be useful to replace the corresponding moment bodies $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$ by convex and compact approximations

$\mathrm{M}^{\mathrm{P}}_{\varepsilon'}$, $\varepsilon' > 0$, that satisfy

$$\mathcal{M}(\mathrm{K})_{\mathrm{P}} \subseteq \mathrm{M}^{\mathrm{P}}_{\varepsilon'} \subseteq \mathrm{N}_{\varepsilon'}(\mathcal{M}(\mathrm{K})_{\mathrm{P}}).$$

Since the linear constraints $\underline{\mathbf{x}}^{\alpha}_{\mathrm{K}} \leq \mathrm{v}_{\alpha} \leq \overline{\mathbf{x}}^{\alpha}_{\mathrm{K}}$ with $\alpha \in \mathrm{P}$ are valid for $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$, one can always add these constraints to the underlying approximate description of $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$. We can therefore assume that $\mathrm{M}^{\mathrm{P}}_{\varepsilon'}$ is a subset of $\mathrm{Box}(\underline{\mathbf{x}}^{\mathrm{P}}_{\mathrm{K}}, \overline{\mathbf{x}}^{\mathrm{P}}_{\mathrm{K}})$. Replacing the separation problems in step (2) of Algorithm 5.1 by the separation problems for $\mathrm{M}^{\mathrm{P}}_{\varepsilon'}$ yields an algorithm that solves

$$\begin{array}{rl} \text{minimize} & \langle\, \mathbf{f}, \mathbf{v}\,\rangle \\ \text{for} & \mathbf{v} \;\in \mathbb{R}^{\mathrm{A}_{\mathcal{F}}} \\ \text{subject to} & \mathbf{v}_{\mathrm{P}} \in \mathrm{M}^{\mathrm{P}}_{\varepsilon'} \text{ for all } \mathrm{P} \in \mathcal{F}. \end{array} \tag{P'-RLX}$$

This algorithm terminates after finitely many iterations. To see this, it suffices to observe that Theorem 5.2 holds for $\mathrm{M}^{\mathrm{P}}_{\varepsilon'}$ in place of $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$, since the proof of Theorem 5.2 only relies on the convexity and compactness of $\mathcal{M}(\mathrm{K})_{\mathrm{P}}$. Similarly, the proof of Theorem 5.3 can be used without any changes to show that the optimal value $\mathrm{f}^{\varepsilon'}$ of (P'-RLX) and the output $\mathbf{v}^{\varepsilon'}(\varepsilon)$ of the algorithm satisfy

$$\langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon)\rangle \leq \mathrm{f}^{\varepsilon'} \leq \|\mathbf{f}\|_{\infty}\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon + \langle \mathbf{f}, \mathbf{v}^{\varepsilon'}(\varepsilon)\rangle. \tag{5.2}$$

Since every feasible point of (P'-RLX) is $\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon$-feasible for (P-RLX'), we have

$$\mathrm{f}^{\star} \leq \|\mathbf{f}\|_{\infty}\mathrm{C}(\mathrm{K}, \mathrm{A}_{\mathcal{F}})\varepsilon' + \mathrm{f}^{\varepsilon'}. \tag{5.3}$$

Combining (5.2) and (5.3) using the triangle inequality yields

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

This line of thought justifies replacing separation problems for chains and shifted chains in step (2) of Algorithm 5.1 by their linear relaxation obtained by applying Proposition 4.20 to the respective separation problem, i.e. (4.14) and (4.15).

5.2 Pattern Relaxation and Conic Programming

We establish what we consider to be a cone program (CP) in standard form:

Definition 5.5 (Conic Standard Form).
Let $\mathcal{K}^1, \ldots, \mathcal{K}^r$ be closed and convex cones, J a finite index set, $\mathbf{b} \in \mathbb{R}^{\mathrm{J}}$ and obj : $\mathcal{K}^1 \times \cdots \times \mathcal{K}^r \to \mathbb{R}$, con : $\mathcal{K}^1 \times \cdots \times \mathcal{K}^r \to \mathbb{R}^{\mathrm{J}}$ linear functions. We say that the cone program

$$\begin{array}{lll} \text{maximize} & obj(\mathbf{x}) & \\ \text{for} & \mathbf{x} & \in \mathcal{K}^1 \times \cdots \times \mathcal{K}^r \\ \text{subject to} & con(\mathbf{x}) = \mathbf{b}. & \end{array} \tag{CP}$$

is in standard form and refer to the cones $\mathcal{K}^i$ that a specific conic programming (CP) solver can handle as its basic cones. △

What a basic cone is depends on the chosen solver. For example among the basic cones that `MOSEK` can handle are: $\mathbb{R}$, $\mathbb{R}_+$, second-order cone, positive semidefinite cones, power cones and their duals. State-of-the-art CP solvers allow to combine different types of basic cones within (CP). These solvers usually employ an interior-point method. For a detailed description of the interior-point method for SDPs see for example [25, Ch. 6]. However, there exists other approaches to solve SDPs such as spectral bundle methods [28].

Next, recall that (D-RLX) was

$$\begin{array}{lll} \text{maximize} & \lambda & \\ \text{for} & \lambda \ \in \mathbb{R}, & \\ & f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}} & \text{for all } \mathrm{P} \in \mathcal{F} \\ \text{subject to} & f - \lambda = \sum_{\mathrm{P} \in \mathcal{F}} f^{\mathrm{P}}. & \end{array} \tag{D-RLX}$$

In order to transform (D-RLX) into standard form only relying on `MOSEK`'s basic cones, we can use the established representations of $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ for the different patterns P:

- For k-variate truncated submonoids we use (approximate) $\mathbb{N}^{\mathrm{k}}_{2\mathrm{d}}$-representations of the cones $\mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})}$ presented throughout Chapter 4 together with Proposition 4.5. Note that, if approximate representations are used, a relaxation of

(D-RLX) is solved and not (D-RLX).

- For a multilinear pattern P, when K is an axis-parallel box, we use Remark 3.6 in order to express $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ using LP constraints.
- For circuits we use Corollary 3.17 and the dual power cone, when $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$. If K is a constrained set $\{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : g^{\mathrm{i}}(\mathbf{x}) \geq 0 \text{ for } \mathrm{i} \in [\mathrm{r}]\}$, where $g^1, \dots, g^{\mathrm{m}} \in \mathbb{R}[x]$ are polynomials, we employ the approximation (C3) of Remark 3.1 to use a circuit P for constraint problems. Note that by doing so, we in fact create a new pattern $\tilde{\mathrm{P}} = \mathrm{P} \cup \bigcup_{\mathrm{i}\in[\mathrm{r}]} \big(\mathrm{supp}(g^{\mathrm{i}}) + \mathrm{P}\big)$ and approximate $\mathcal{P}(\mathrm{K})_{\tilde{\mathrm{P}}}$.

5.3 Graph Based Pattern Selection

In order for a pattern family $\mathcal{F}$ to induce a pattern relaxation of (POP) it is necessary that $\mathrm{A} \subseteq \mathrm{A}_{\mathcal{F}}$. This minimal requirement allows for a wide variety of different pattern families. Throughout Chapters 3 and 4 we have already encountered different families:

- the families depicted in the figures that were used to illustrate the pattern types
- $\{\mathrm{P}^\alpha : \alpha \in \mathrm{A}\}$ and $\{\mathrm{P} : \mathrm{P} \in \mathcal{F}^\alpha \text{ for } \alpha \in \mathrm{A}\}$ corresponding to pattern relaxations (3.5) and (3.6)
- the expression tree approach, discussed in Section 3.3, induces a pattern family
- the pattern family in Corollary 3.22 contains all binomial square patterns in $\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}$ for the relaxation of $\mathcal{P}(\mathbb{R}^{\mathrm{n}})_{\mathbb{N}^{\mathrm{n}}_{2\mathrm{d}}}$.

Furthermore, in [58] and [39] heuristic methods are used to generate families of circuit patterns and in [47] an iterative method is proposed that augments a given family of circuits after solving both (P-RLX) and (D-RLX). These pattern selection strategies do not actively exploit the monomial structure encoded in A. Furthermore, none of these families take into account that each pattern P comes with different computational costs for enforcing the constraint $f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}}$ in (D-RLX). Motivated by this we present a pattern selection routine that takes both into account: the structure of A and the cost per pattern P.

This leads to the overdue question: What do we mean by computational cost? The answer to this is user dependent as this might be time complexity or memory complexity. Whatever one chooses, we assume that we can assign to each pattern P computational cost $\mathrm{co}(\mathrm{P}) \geq 0$ such that the computational cost of solving (D-RLX) is proportional to $\sum_{\mathrm{P}\in\mathcal{F}} \mathrm{co}(\mathrm{P})$. This assumption is reasonable as the constraints $f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}}$ of (D-RLX) can be enforced independently for each $\mathrm{P} \in \mathcal{F}$ such that

(D-RLX) has a block structure. This structure in turn can be exploited by conic programming solvers.

As there exists a potentially infinite number of patterns $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ to choose from, we first limit this choice by using a pattern family $\mathcal{A}$. This family $\mathcal{A}$ shall contain all patterns that we deem to be useful for a pattern relaxation of A. It is our goal to select a pattern family $\mathcal{F} \subseteq \mathcal{A}$ that has the following features:

(F1) $\mathrm{A} \subseteq \mathrm{A}_{\mathcal{F}}$,

(F2) the original exponents in A are connected among each other (directly or indirectly) and

(F3) the computational cost $\sum_{\mathrm{P}\in\mathcal{F}} \mathrm{co}(\mathrm{P})$ corresponding to $\mathcal{F}$ are minimal among all families that satisfy (F1) and (F2).

Next we borrow some notation from combinatorial optimization, see for example [33]: A graph G is a tuple $\mathrm{G} = (\mathrm{No}\,(\mathrm{G})\,, \mathrm{Ed}\,(\mathrm{G}))$, where $\mathrm{No}\,(\mathrm{G})$ is a finite and nonempty set and $\mathrm{Ed}\,(\mathrm{G})$ is finite subset of $\{\{\mathrm{w}, \mathrm{u}\} \subseteq \mathrm{No}\,(\mathrm{G}) : \mathrm{w} \neq \mathrm{u}\}$. We call the elements of $\mathrm{No}\,(\mathrm{D})$ nodes and the elements of $\mathrm{Ed}\,(\mathrm{D})$ edges. A subgraph of G is a graph $\mathrm{S} = (\mathrm{No}\,(\mathrm{S})\,, \mathrm{Ed}\,(\mathrm{S}))$ with $\mathrm{No}\,(\mathrm{S}) \subseteq \mathrm{No}\,(\mathrm{G})$ and $\mathrm{Ed}\,(\mathrm{S}) \subseteq \mathrm{Ed}\,(\mathrm{G})$. We write $\mathrm{S} \subseteq \mathrm{G}$ to indicate that S is subgraph of G. Let $\mathrm{a}_1, \mathrm{a}_{\mathrm{k}+1} \in \mathrm{No}(\mathrm{G})$ with $\mathrm{a}_1 \neq \mathrm{a}_{\mathrm{k}+1}$. A a_1-$\mathrm{a}_{\mathrm{k}+1}$-path in G is a subgraph

$$\mathrm{H} = (\{\mathrm{a}_1, \ldots, \mathrm{a}_{\mathrm{k}+1}\}\,, \{\{\mathrm{a}_1, \mathrm{a}_2\}, \ldots, \{\mathrm{a}_{\mathrm{k}}, \mathrm{a}_{\mathrm{k}+1}\}\}) \subseteq \mathrm{G}$$

with $\mathrm{a}_{\mathrm{i}} \neq \mathrm{a}_{\mathrm{j}}$ for all $\mathrm{i} \neq \mathrm{j}$. The length of H is the cardinality of Ed(H), i.e. $\#\,\mathrm{Ed}(\mathrm{H}) = \mathrm{k}$. By abuse of notation we refer to Ed(H) as a_1-a_{k}-path, since we can always retrieve No(H) from Ed(H). We assign weights to the nodes using a function $\mathrm{we} : \mathrm{No}(\mathrm{G}) \to \mathbb{R}$. For $\mathrm{N} \subseteq \mathrm{No}(\mathrm{G})$ and $\mathrm{S} \subseteq \mathrm{G}$ we write $\mathrm{we}(\mathrm{N}) := \sum_{\mathrm{a}\in\mathrm{N}} \mathrm{we}(\mathrm{a})$ and $\mathrm{we}(\mathrm{S}) := \mathrm{we}(\mathrm{No}(\mathrm{S}))$. We can now define a node-weighted graph $\mathrm{G}\,(\mathcal{A})$ for the pattern selection routine. The nodes and edges of this graph are

$$\begin{aligned}\mathrm{No}(\mathrm{G}\,(\mathcal{A})) :=& \mathrm{A} \cup \mathcal{A}\\ \mathrm{Ed}(\mathrm{G}\,(\mathcal{A})) :=& \Big\{\big\{\alpha, \mathrm{P}\big\} : \alpha \in \mathrm{A}, \mathrm{P} \in \mathcal{A} \text{ with } \alpha \in \mathrm{P}\Big\} \cup\\ & \Big\{\{\mathrm{P}, \tilde{\mathrm{P}}\} \subseteq \mathcal{A} : \tilde{\mathrm{P}} \neq \mathrm{P} \text{ and } \tilde{\mathrm{P}} \cap \mathrm{P} \backslash \{\mathbf{0}\} \neq \emptyset\Big\}\end{aligned}$$

and we define the weight of an node $\mathrm{a} \in \mathrm{No}\,(\mathrm{G}\,(\mathcal{A}))$ by

$$\mathrm{we}(\mathrm{a}) := \begin{cases}\mathrm{co}(\mathrm{a}), & \text{if } \mathrm{a} \in \mathcal{A},\\ 0, & \text{else.}\end{cases}$$

Let $\alpha, \beta \in \mathrm{A}$ with $\alpha \neq \beta$ and

$$\mathrm{H} = \left\{\left\{\alpha, \mathrm{P}_1\right\}, \left\{\mathrm{P}_1, \mathrm{P}_2, \right\}, \ldots, \left\{\mathrm{P}_{\mathrm{k}-1}, \mathrm{P}_{\mathrm{k}}\right\}, \left\{\mathrm{P}_{\mathrm{k}}, \beta\right\}\right\}.$$

be α-β-path in $\mathrm{G}(\mathcal{A})$. The path H induces a pattern family $\{\mathrm{P}_{\mathrm{i}}\}_{\mathrm{i}\in[\mathrm{k}]}$ that connects $\alpha \in \mathrm{P}_1$ and $\beta \in \mathrm{P}_{\mathrm{k}}$. Thus, an optimal solution $\mathrm{S}^{\mathrm{opt}}$ of

$$\min\left\{\mathrm{we}\left(\mathrm{S}\right) : \mathrm{S} \subseteq \mathrm{G}\left(\mathcal{A}\right) \text{ and there exists } \alpha\text{-}\beta\text{-path in S for all } \alpha, \beta \in \mathrm{A}\right\} \tag{5.4}$$

induces a pattern family $\mathcal{F} = \mathrm{No}(\mathrm{S}^{\mathrm{opt}})\backslash\mathrm{A}$ with costs

$$\mathrm{we}\left(\mathrm{S}^{\mathrm{opt}}\right) = \sum_{\mathrm{i}\in[\mathrm{k}]} \mathrm{co}(\mathrm{P}_{\mathrm{i}}).$$

This family has, due to the construction of the graph $\mathrm{G}(\mathcal{A})$, the features (F1), (F2) and (F3). The problem (5.4) is known as node-weighted Steiner tree problem [33, Ch. 20]. This well investigated problem [52] has been implemented in the application `SCIP-Jack` [23] of `SCIP` [24], which we used for our implementation. Before we address in Example 5.6 the question how to choose $\mathcal{A}$ and the cost function for multilinear patterns, we note that this is merely one possible construction of such a graph. A natural extension could be assigning capacities to the edges $\{\mathrm{P}, \tilde{\mathrm{P}}\}$ in $\mathrm{G}(\mathcal{A})$ that reflect the cardinality of $\tilde{\mathrm{P}} \cap \mathrm{P}\backslash\{\mathbf{0}\}$. Another, different, approach is to use a hypergraph with nodes in $\mathbb{N}^{\mathrm{n}}$ that correspond to exponents and edges $\mathrm{P} \subseteq \mathbb{N}^{\mathrm{n}}$ correspond to patterns.

Example 5.6.
For a given set of exponents $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ *we define* $\mathcal{A}$ *using* $\mathrm{I} := \{\mathbf{1}, \mathbf{e}^1, \ldots, \mathbf{e}^{\mathrm{n}}, \mathbf{0}\}$ *and*

$$\bar{\mathrm{A}} := \{\alpha \in \mathbb{N}^{\mathrm{n}} : \text{for all } \mathrm{i} \in [\mathrm{n}] \text{ there exists } \beta \in \mathrm{A} \text{ with } \alpha_{\mathrm{i}} = \beta_{\mathrm{i}}\}$$

$\mathcal{A} := \{\mathrm{ML}(\alpha, \mathrm{I}) : \alpha \in \bar{\mathrm{A}} \text{ and } \mathrm{ML}(\alpha, \mathrm{I}) \geq 3\}$*. The auxiliary set* $\bar{\mathrm{A}}$ *contains* A *and yields 'orthogonal structures' that are well suited for exploitation with multilinear patterns from* $\mathcal{A}$*. The idea is that if* A *is not ill-structured and sparse then the cardinality of* $\mathcal{A}$ *will be moderate. In turn, this keeps the number of nodes and edges of* $\mathrm{G}(\mathcal{A})$ *relatively small compared to a choice a of* $\mathcal{A}$ *such as* $\mathcal{A} = \{\mathrm{ML}(\alpha, \mathrm{I}) : \alpha_{\mathrm{i}} \leq \max_{\beta\in\mathrm{A}} \beta_{\mathrm{i}} \text{ for all } \mathrm{i} \in [\mathrm{n}]\}$ *– which is desirable as the Steiner tree problem is* $\mathcal{NP}$*-hard [52]. The next figure illustrates these orthogonal structures on* $\mathrm{A} = \{(2, 3), (2, 4), (5, 5)\}$.

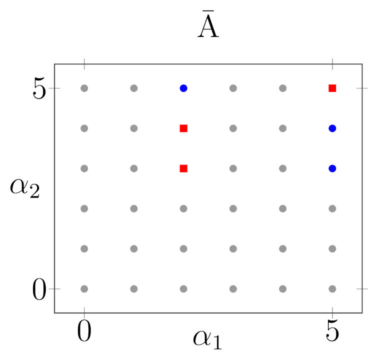

Figure 5.1: As usual, we depict the original exponents in $\mathrm{A} = \{(2,3),(2,4),(5,5)\}$ by red squares and the auxiliary exponents in $\bar{\mathrm{A}}\backslash\mathrm{A}$ by blue dots.

For simplicity, we use $\mathrm{co}(\mathrm{P}) := \#\mathrm{P}$ *as a proxy for the computational cost of a multilinear pattern* P*. Applying this procedure to the exponent sets* A_{ex} *and* A_2 *of the previous chapters yields the graphs and minimal node-weighted Steiner trees depicted in Figure 5.2. The graphs are already fairly large. In order to further reduce their sizes an option is to exploit the symmetries that they admit. However, as the pattern families corresponding to the Steiner trees, shown in Figure 5.3, already have the desired properties, we leave it at this. In particular, the right subplot of Figure 5.3 shows a pattern family for which all exponents in* A_2 *are indirectly connected – an improvement to the family shown in Figure 3.1.* $\triangle$

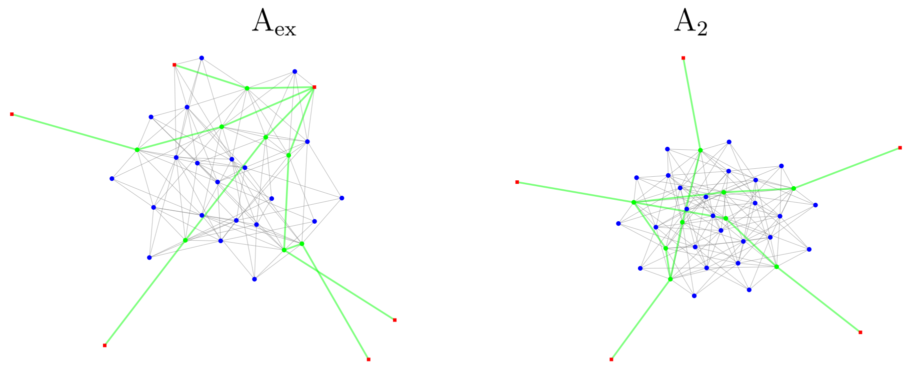

Figure 5.2: Two Graphs $G(\mathcal{A})$ are shown for $\mathcal{A}$ constructed from A_{ex} and A_2 as described in Example 5.6. The red squares corresponds to the nodes A_{ex} and colored dots to nodes $\mathcal{A}$. The green dots and edges together with the red squares depict a minimal node-weighted Steiner tree as computed by `SCIP-Jack`.

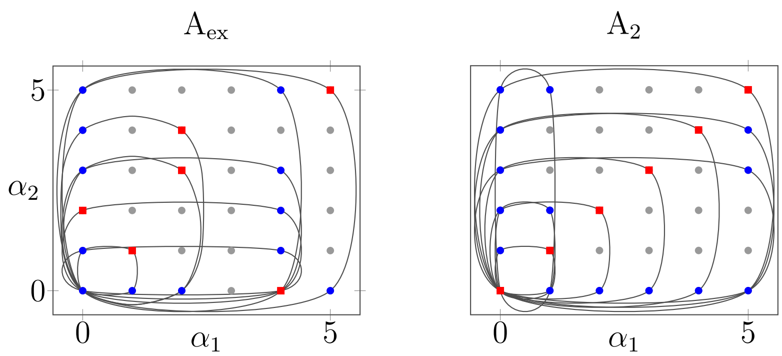

Figure 5.3: Visualization of the multilinear families that correspond to minimal node-weighted Steiner trees shown in Figure 5.2. The format is as described in Plot Set Up 3.2.

Chapter 6

Numerical evaluation

Finding an unbiased and objective setting to compare the advantages and disadvantages of convex relaxations for POPs is not trivial, as models, their purpose, and methods are usually closely linked to one another. We decided to use two different implementations of the pattern relaxation, namely `MONORELS` and `DUAL-MONORELS`. The first implementation, `MONORELS`, is a prototype of Algorithm 5.1 that relies purely on linear programming to compute solutions of (P-RLX') for different pattern families. The second implementation, `DUAL-MONORELS`, translates (D-RLX) into conic standard form and then computes its optimal value using `MOSEK`. `DUAL-MONORELS` allows to combine multilinear patterns, all variants of truncated submonoids (including chains and binomial square patterns) and circuits. The solutions are used to approximate the size of the relaxations. As measure for the size, the width of the relaxations in different directions is used. We use the width of the relaxations in different directions as a measure for the size. Different pattern relaxations are compared on a new benchmark library of random POP instances among another and to results from different solvers. This demonstrates how different pattern families influence the quality of the pattern relaxations. As reference we use the widths computed by `BARON`, the arguably leading solver for nonconvex optimization, and the widths computed by `TSSOS`, a sparsity exploiting SDP based approach.

We start by describing implementation and comparison details, before numerical results for different classes of instances are discussed.

6.1 Implementation Details

Four different solvers are run for the numerical evaluation on a compute server with 4 Intel(R) Xeon(R) Gold 6138 CPUs with 20 cores of 2 threads and 1 TB RAM each under Ubuntu 18.04.4. Each solver-instance pair was assigned to one such job, i.e. the solvers themselves did not use the parallel structure. In order to distribute the solver-instance pairs to the 80 cores we used [65]. The following versions were used `Matlab` 9.6.0.1174912 (R2019a) Update 5 [41], `MOSEK` 9.2.32 [3], `JULIA` 1.5.2 [13], `TSSOS` version 1.00 [67] and `BARON` 1.8.9 [66]. All reported run times are real times. The codes of MONORELS and DUAL-MONORELS are implemented and run in Matlab. MONORELS consists of roughly 3000 lines for the main algorithm, relaxations of different patterns, utilities, and unit tests. As a subsolver for the linear programs the dual simplex algorithm of MOSEK's linprog was used (without warmstarting) for the master and the separation problems for chains and shifted chains arising from Proposition 4.20 with $\varepsilon = 10^{-4}$. For the covering $\mathcal{I}$ we chose an interval decomposition of 9 equidistant intervals. The cutting-plane algorithm of MONORELS was timed with MATLAB's tic and toc commands. The code of DUAL-MONORELS consists of roughly 3500 lines of code and uses MOSEK to solve the cone programs. The reported time is the termination time obtained from MOSEK. BARON [66] was called from Matlab with default settings. BARON currently only returns the CPU time, when its Matlab interface is used. Hence, we time a BARON call with MATLAB's tic and toc commands[1]. TSSOS is a JULIA package that allows to exploit correlative sparsity and term sparsity simultaneously. We called the first level of the hierarchy by running the command cs_tssos_first with settings order $= \lceil \frac{\deg(f)}{2} \rceil$ and TS="MD". With these settings the computation time of TSSOS is typically the fastest, but may yield looser bounds than other settings. TSSOS does not report the time of the solution process. Thus, we first piped the output from the SDP solver MOSEK, that TSSOS uses, to a text file. After that we read the termination time of MOSEK from the text file. Since only two decimal places are obtained this way, the time we report is only a proxy of the time of TSSOS's actual MOSEK call. The code for of MONORELS and DUAL-MONORELS as well as the code used for calling BARON and TSSOS is on the CD that is included in the back cover.

6.2 Setup of Numerical Comparisons

As an indicator for the tightness of relaxations we approximate the size of feasible sets by their width. For a given finite and nonempty set $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ and a vector $\mathbf{f} \in \mathbb{R}^{\mathrm{A}}$ we define the width function $\omega_{\mathcal{M}(\mathrm{K})_{\mathrm{A}}}(\mathbf{f})$ of $\mathcal{M}(\mathrm{K})_{\mathrm{A}}$ in direction $\mathbf{f}$ as

$$\omega_{\mathcal{M}(\mathrm{K})_{\mathrm{A}}}(\mathbf{f}) = \max_{\mathbf{x} \in \mathrm{K}} f(\mathbf{x}) - \min_{\mathbf{x} \in \mathrm{K}} f(\mathbf{x}). \tag{6.1}$$

Replacing K by a relaxation based on a pattern family $\mathcal{F}$ one obtains an upper bound on the value of $\omega_{\mathcal{M}(\mathrm{K})_{\mathrm{A}}}(\mathbf{f})$, denoted by $\omega(\mathcal{F}, \mathcal{M}(\mathrm{K})_{\mathrm{A}}, \mathbf{f})$. The evaluation requires solving two instances of (P-RLX') for every pattern of interest, using the objective functions $\langle -\mathbf{f}, \mathbf{v} \rangle$ and $\langle \mathbf{f}, \mathbf{v} \rangle$, respectively. To normalize the values $\omega_{\mathcal{M}(\mathrm{K})_{\mathrm{A}}}(\mathbf{f})$ and $\omega(\mathcal{F}, \mathcal{M}(\mathrm{K})_{\mathrm{A}}, \mathbf{f})$ to the range $[0, 1]$, we divide by the width function obtained for the

[1]This method was suggested with the support of BARON.

(trivial) relaxation using the singletons-only pattern $\mathcal{F}_{\mathrm{A}}^{\mathrm{sgl}} = \{\{\alpha\} : \alpha \in \mathrm{A}\}$, i.e.,

$$\nu(\mathcal{P}, \mathrm{A}, \mathbf{f}) := \frac{\omega(\mathcal{P}, \mathcal{M}(\mathrm{K})_{\mathrm{A}}, \mathbf{f})}{\omega(\mathcal{P}_{\mathrm{A}}^{\mathrm{sgl}}, \mathcal{M}(\mathrm{K})_{\mathrm{A}}, \mathbf{f})} \quad \text{and} \quad \nu_{\mathrm{A}}(\mathbf{f}) := \frac{\omega_{\mathcal{M}(\mathrm{K})_{\mathrm{A}}}(\mathbf{f})}{\omega(\mathcal{P}_{\mathrm{A}}^{\mathrm{sgl}}, \mathcal{M}(\mathrm{K})_{\mathrm{A}}, \mathbf{f})}. \tag{6.2}$$

Tables 6.1 and 6.2 list the methods and patterns that were used for the numerical results. Method (B) gives an approximation of the reference solution, albeit at a high computational cost. (R) can be seen as the current state-of-the-art for a relaxation within a divide-and-conquer approach. (CS) is, to our knowledge, the current state-of-the-art of sparsity exploiting SDP approaches. Our approach allows to compare the new relaxation strategies (M), (C), (MC), (H) and (T) with respect to the width function.

6.3 Test Instances

Our set of test instances consists of 13 nonempty and finite exponent sets $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ which were classified into four types: specially structured adversary sets, dense sets, sparse sets, and the example A_{ex} from above. They are explained in the next subsection. For each exponent set we chose $\mathrm{K} = [0,1]^{\mathrm{n}}$ and 20 (uniformly distributed) random coefficient vectors $\mathbf{f}^1, \ldots, \mathbf{f}^{20} \in [-1,1]^{\mathrm{A}}$. The instances were a priori filtered to avoid trivial problems. If `BARON` did terminate the minimization or the maximization task in (6.1) within the CPU time limit of 1000 seconds[2], the instance was replaced. Therefore the corresponding mean times for (B) are always at least 1000 seconds.

The motivation was not to show that our new relaxations are always better than the ones within `BARON` or `TSSOS`, but that there are difficult instances for which an improvement is possible.

6.4 Numerical Results

In this subsection we describe the different exponent sets and show numerical results for the different methods from Tables 6.1 and 6.2.

Box Plot Set Up 6.1.

Figures 6.1 to 6.5 and 6.7 show box plots of our numerical findings. The box plots visualize the distributions (20 random vectors $\mathbf{f}^{\mathrm{i}}$) of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 computed with `BARON`, `MONORELS`

[2]To avoid confusion: `BARON` reports the CPU time but not the real time.

DUAL-MONORELS and TSSOS. The title of a subplot corresponds to the exponent set A*. Below the method (see Tables 6.1 and 6.2) the rounded mean time in seconds is shown for the respective method. The box plots for* (6.2) *computed by MONORELS and DUAL-MONORELS are indistinguishable to the human eye. Thus, we only show the box plots for the widths and computed by DUAL-MONORELS when the methods (C), (S), (M), (MC), (H) and (T) are used. The time right below these methods is the one reported by DUAL-MONORELS. Additionally, if computed, the mean time for the bounds computed by MONORELS can be found beneath the respective time from DUAL-MONORELS. The box borders are the* 1/4 *and the* 3/4*-quantiles. The lower whisker is the smallest data value which is larger than the lower quartile* -1.5 *times the interquartile range and the upper whisker accordingly.* △

Label	Description
(B)	Reference solution: To approximate $\omega_{\mathcal{M}(\mathrm{K})_\mathrm{A}}(\mathbf{f})$ we use the best upper bound for $\max_{\mathbf{x}\in\mathrm{K}} f(\mathbf{x})$ and the best lower bound for $\min_{\mathbf{x}\in\mathrm{K}} f(\mathbf{x})$ that BARON returns within a CPU time limit of 1000 seconds each.
(R)	Root node relaxation of the BARON solver.
(CS)	Reference solution obtained from the sparsity exploiting solver TSSOS.
(SOS)	Self-implemented SOS relaxation of the lowest hierarchy level that does not exploit sparsity (see Section 2.1).
(M)	Relaxation based on the multilinear pattern $\mathcal{F}^{\mathrm{m}}_{\mathrm{A}}$, which consists of the inclusion-maximal elements of $\{\mathrm{ML}(\alpha, \{0,1\}^{\mathrm{n}}) : \alpha \in \mathrm{A}\backslash\{0\}\}$.
(S)	Relaxation based on a family of shifted chains $\mathcal{F}^{\mathrm{s}}_{\mathrm{A}}$, which consists of the inclusion-maximal elements of $$\{\xi + \mathrm{CH}(\mathbf{e}^{\mathrm{i}}, \mathrm{d}) : \mathrm{d} \in 2\mathbb{N}\backslash\{0\}, \xi \in \mathbb{N}^{\mathrm{n}}, \mathrm{i} \in [\mathrm{n}]\},$$ that satisfy $\#(\xi + \mathrm{CH}(\gamma, \mathrm{d})) \cap \mathrm{A} \geq 2$ and $$\#(\xi + \mathrm{CH}(\gamma, \mathrm{d})) \cap \mathrm{A} > \#(\xi + \mathrm{CH}(\gamma, \mathrm{d}-1)) \cap \mathrm{A}.$$ The latter conditions ensure that each shifted chain contains at least two exponents from A and that we cannot include more exponents from A if we choose a bigger d.

Table 6.1: Methods and relaxations to be compared in Figures 6.1 to 6.5 and 6.7.

Label	Description
(C)	Relaxation based on a family of chains $\mathcal{F}^{c}_{A}$, which consists of the inclusion-maximal elements of $$\{CH(\gamma, d) : d \in 2\mathbb{N}\setminus\{0\}, \gamma \in \mathbb{N}^n\},$$ that satisfy $\#\, CH(\gamma, d) \cap A \geq 2$ and $$\#\, CH(\gamma, d) \cap A > \#\, CH(\gamma, d-1) \cap A.$$
(MC)	Relaxation based on a family of multilinear patterns, chains and shifted chains, $$\mathcal{F}^{mc}_{A} := \mathcal{F}^{m}_{A} \cup \mathcal{F}^{c}_{A_1} \cup \mathcal{F}^{s}_{A_2}$$ where $A_1 := A_{\mathcal{F}^{m}_{A}}$, $A_2 := A_{\mathcal{F}^{m}_{A} \cup \mathcal{F}^{c}_{A_1}}$.
(H)	Let $d(A) := \max(\{\alpha_i : \alpha \in A, i \in [n]\})$ and $\Gamma := \{\mathbf{1}, \mathbf{e}^1, \ldots, \mathbf{e}^n\}$. A relaxation based on the family $$\mathcal{F}^{h}_{A} := \{CH(\gamma, d(A)) : \gamma \in \Gamma\} \cup \mathcal{F}^{m}_{\{CH(\gamma,d(A)):\gamma\in\Gamma\}} \cup \mathcal{F}^{m}_{A},$$ which uses $n+1$ chains that are linked by $d(A)$ multilinear patterns to strengthen $\mathcal{F}^{m}_{A}$.
(T)	Let $d_1 := 2 \cdot \lceil \deg(A)/2 \rceil$, $d_2 := 2 \cdot \lceil \deg(A)/4 \rceil$ and $\Gamma := (2\mathbf{e}^1, \ldots, 2\mathbf{e}^n)$. A relaxation based on the family $\mathcal{F}^{t}_{A}$, which consists of the inclusion-maximal elements of $$\{TS((\mathbf{e}^i)_{i\in supp(\alpha)}, d_1) : \alpha \in A \setminus TS(\Gamma, d_2)\} \cup \{TS(\Gamma, d_2)\}.$$ Here, $(\mathbf{e}^i)_{i\in supp(\alpha)}$ is a matrix with columns $\mathbf{e}^i, i \in supp(\alpha)$. The family $\mathcal{F}^{t}_{A}$ uses k-variate truncated submonoids with $k \leq d_1$ to cover the exponents in A and connects these using one n-variate truncated submonoid.

Table 6.2: Methods and relaxations to be compared in Figures 6.1 to 6.5 and 6.7.

Adversary Exponent Sets

If a pattern family yields only poor connective properties for an exponent set, we consider this set to be an adversary exponent set for this family. In plot A_2 from Figure 3.1, for example, we see that the sparse family $\mathcal{F}^{\mathrm{m}}_{\mathrm{CH}(\mathbf{1},5)}$ of multilinear patterns connects none of the original exponents. Hence, chain shaped exponent sets are natural adversaries for relaxations that only use multilinear patterns. As a result, Figure 6.1 shows that the bounds using $\mathcal{F}^{\mathrm{m}}_{\mathrm{CH}(\gamma,\mathrm{d})}$ (M) coincide with the bounds obtained by the weakest pattern family $\mathcal{F}^{\mathrm{sgl}}_{\mathrm{CH}(\gamma,\mathrm{d})}$. On the other hand, it is not surprising that the bounds obtained by using one chain (C) match the reference solution (B). Furthermore, the sparsity exploiting solver `TSSOS` fails to terminate for any of the 20 instances with exponent set $\mathrm{CH}(\mathbf{1}^4, 10)$. We suspect the reason for this is that $\mathrm{CH}(\mathbf{1}^4, 10)$ does not yield any term or chordal sparsity structures that can be exploited. Thus, `TSSOS` solves a regular moment relaxation for $\mathrm{n} = 4$ and $\deg(\mathrm{CH}(\mathbf{1}^4, 10)) = 40$. This involves computing a SDP that includes a $\binom{24}{4} \times \binom{24}{4}$ PSD matrix. Note that $\binom{24}{4} = 10626$.

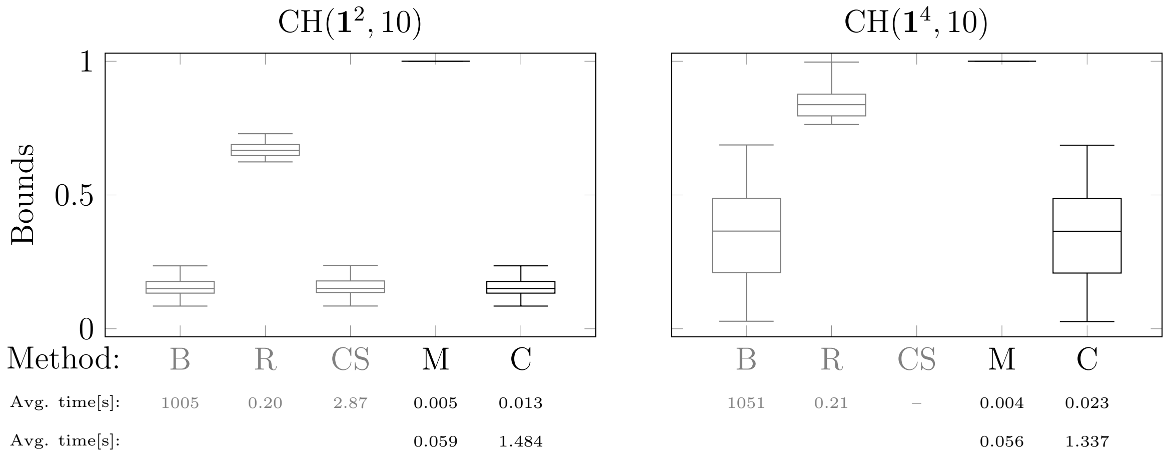

Figure 6.1: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 and for adversary exponent sets.

Another adversary exponent set for multilinear patterns is $\mathrm{C}(\mathrm{n}, \mathrm{d}) := \mathrm{CH}(\mathbf{1}, \mathrm{d}) \cup \mathrm{CH}(\mathbf{e}^1, \mathrm{d}) \cup \cdots \cup \mathrm{CH}(\mathbf{e}^{\mathrm{n}}, \mathrm{d})$. It can be covered sparsely by d multilinear patterns using the family $\mathcal{F}^{\mathrm{m}}_{\mathrm{C}(\mathrm{n},\mathrm{d})}$. Each pattern of $\mathcal{F}^{\mathrm{m}}_{\mathrm{C}(\mathrm{n},\mathrm{d})}$ connects $\mathrm{n}+1$ original exponents, but establishes no connection between monomials from different patterns. That is because two patterns $\mathrm{P}, \mathrm{P}' \in \mathcal{F}^{\mathrm{m}}_{\mathrm{C}(\mathrm{n},\mathrm{d})}$ with $\mathrm{P} \neq \mathrm{P}'$ satisfy $\mathrm{P} \cap \mathrm{P}' = \{\mathbf{0}\}$. The poor connective properties of $\mathcal{F}^{\mathrm{m}}_{\mathrm{C}(\mathrm{n},\mathrm{d})}$ explain their poor performance, see (M) in Figure 6.2. By additionally using $\mathrm{n}+1$ chains to connect the d multilinear patterns, the family

$\mathcal{F}^{\mathrm{h}}_{\mathrm{C(n,d)}}$ exploits the structure of $\mathrm{C(n,d)}$. As a consequence, the bounds computed with (H) and (B) are indistinguishable in Figure 6.2. Again, `TSSOS` fails to terminate for any of the instances with exponent set $\mathrm{C}(4, 10)$ – most likely, because no term or chordal sparsity structures can be found.

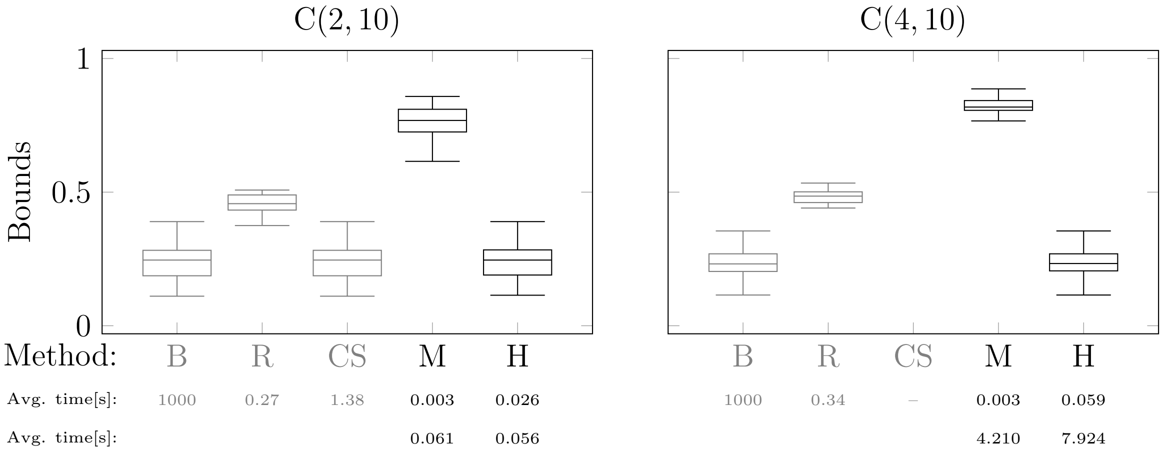

Figure 6.2: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 and for adversary exponent sets.

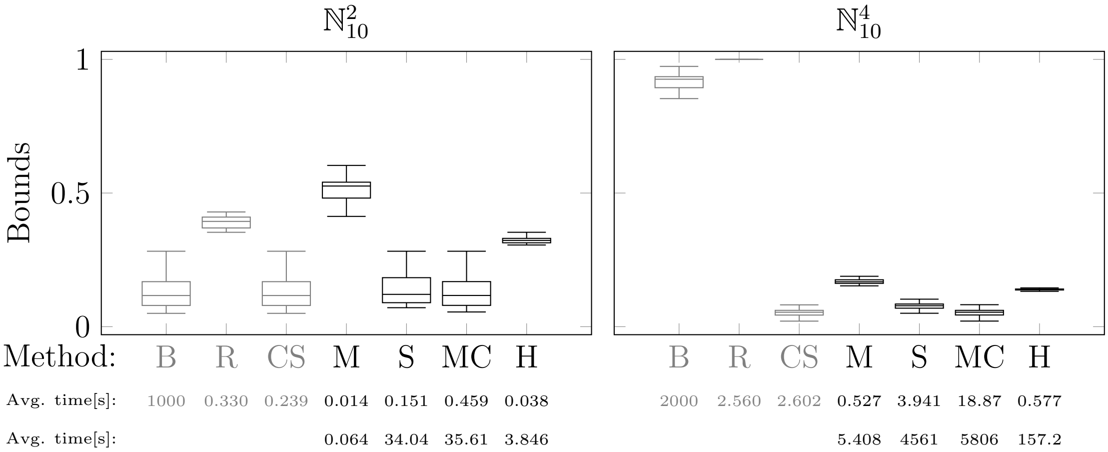

Figure 6.3: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 and for dense exponent sets.

Dense Exponent Sets

We consider dense exponent sets $A = \mathbb{N}^n_d$ for $n \in \{2, 4\}$ and $d = 10$. The pattern families shown in Figure 6.3 perform reasonably well, probably due to their connectivity properties. Furthermore, we see that the multilinear patterns (M) perform for $n = 4$ drastically better than (R). This might be because the multilinear patterns $ML(\alpha, \{0, 1\}^n)$ used in $\mathcal{F}^m_{\mathbb{N}^4_{10}}$ are bigger than the ones `BARON` uses, leading to more connections between monomial variables.

Sparse Exponent Sets

We use randomly generated sparse exponent sets $A = S(n, d)$ to test pattern families that do not assume any structure of A. $S(n, d)$ is generated by randomly picking $\lceil \binom{n+d}{d} \rceil$ exponents via `Matlab's randperm` from $\mathbb{N}^n_d$.

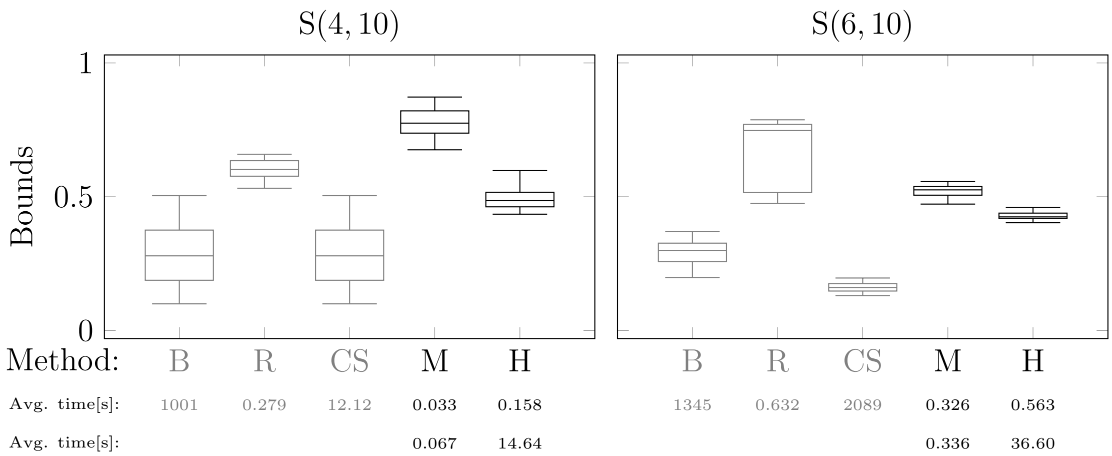

Figure 6.4: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 and for sparse exponent sets of degree 10.

Figure 6.4 column (M) shows that $\mathcal{F}^m_{S(n,d)}$ does not perform particularly well. Column (H) shows that additionally enforcing indirect connections between moment variables via $n + 1$ chains and d multilinear patterns in $\mathcal{F}^h_{S(n,d)}$ results in tighter bounds.

Figure 6.5 shows the distribution of the width for sparse instances with a high number of variables $n = 25, 30, 35, 40$ and low degree $d = 4$. These instances are infeasible for relaxations that do not exploit sparsity of $S(n, d)$ if $n \geq 35$. Thus, it was impossible to compute the width with (SOS) if $n \geq 35$. Interestingly, the bounds computed by `TSSOS` are worse than the ones computed with the pattern family $\mathcal{F}^{sgl}_{S(n,4)}$. It might be that using different settings for `TSSOS` yields better bounds. However, this would also

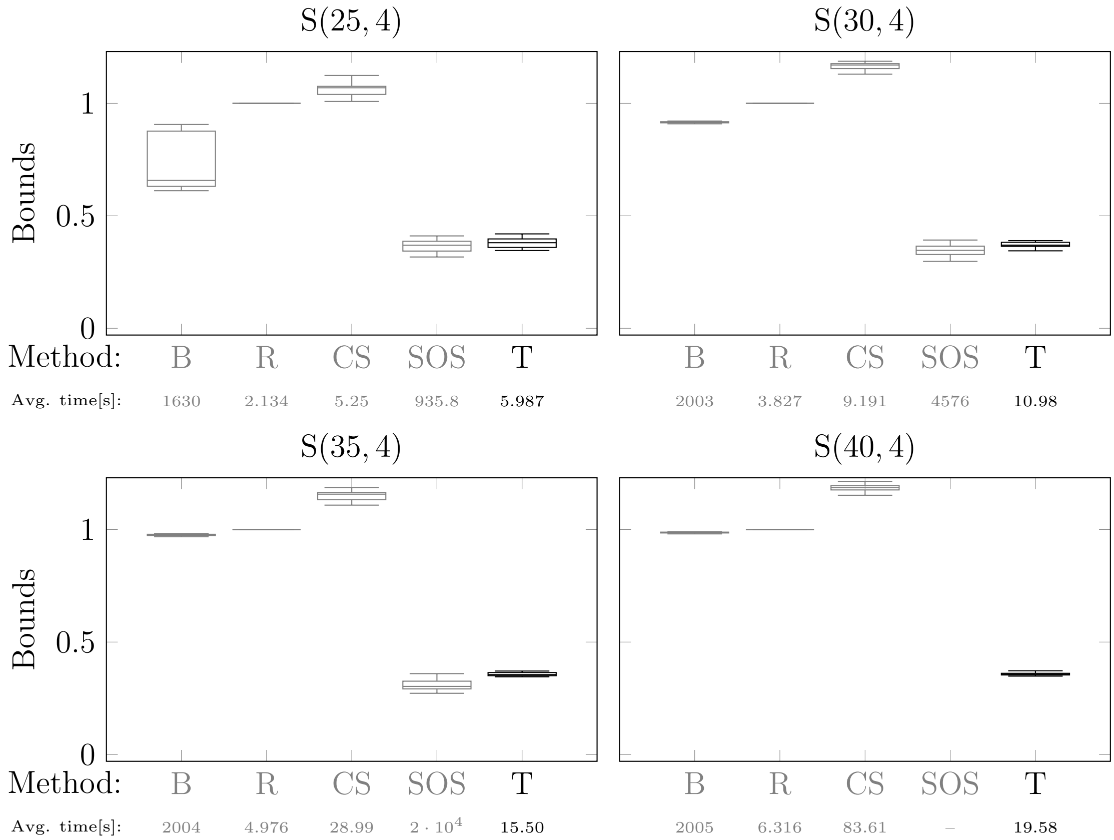

Figure 6.5: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 and for sparse exponent sets with degree 4. The correct average time (rounded to integers) for (SOS) and $\mathrm{S}(35, 4)$ is 20246s, which rounds to more than 5.5h.

result in higher computation times. The pattern strategy (T) yields for all tested n nontrivial bounds. Note that for $\mathrm{n} = 25, 30, 35$ these bounds seem to be reasonably tight, when compared to (SOS), but for a fraction of the computation time of (SOS). For $\mathrm{n} = 40$ (T) outperforms all other tested relaxation methods, that terminated,[3] in terms of the quality of the bounds and is only bested by the root node relaxation (R) of `BARON` in terms of time. We want to point out that we were able to compute nontrivial bounds for instances with exponent sets $\mathrm{S}(80, 4)$. For these exponent sets the computation of the width in one direction usually takes between 6-7 minutes. The reason for the good performance in terms of computation time of (T) can be traced back to the relatively small size of the biggest involved $\mathrm{r} \times \mathrm{r}$ PSD matrices in the relaxation of (D-RLX). That is for $\mathcal{F}^{\mathrm{t}}_{\mathrm{S(n,2d)}} = \{\mathrm{TS}(\{\mathbf{e}^{\mathrm{i}}\}_{\mathrm{i} \in \mathrm{supp}(\alpha)}, 2\mathrm{d}) : \alpha \in$

[3] To be more precise: that terminated for all 20 coefficient vectors within one week.

$S(n, 2d) \setminus TS(\Gamma, 2d)\} \cup \{TS(\Gamma, d)\}$

$$r \leq \max\left\{\binom{d + \min\{n, 2d\}}{\min\{n, 2d\}}, \binom{\lceil \frac{d}{2} \rceil + n}{n}\right\}.$$

This boils for $2d = 4$ and $n \geq 4$ down to $r \leq \max\left\{\binom{2+4}{4}, \binom{1+n}{n}\right\}$.

Custom Strategies

A customized pattern family $\mathcal{P}$ for a given exponent set A allows to trade off computational cost versus tightness of the relaxation. Figure 6.6 shows three example pattern families customized for A_{ex} from Plot Set Up 3.2.

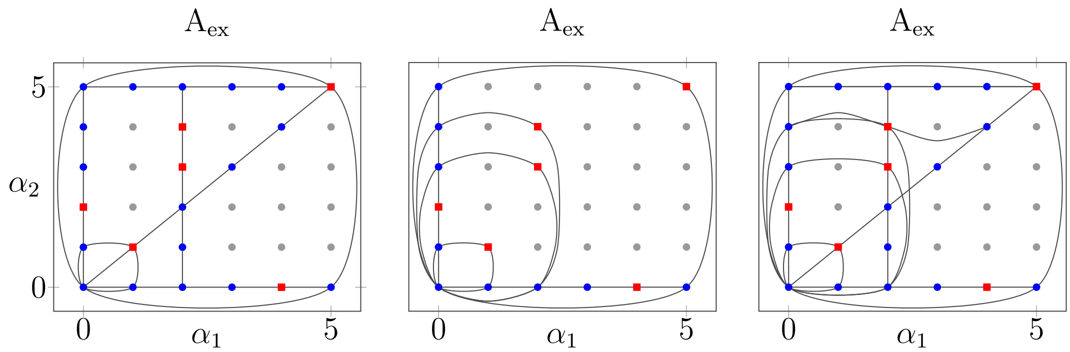

Figure 6.6: Visualization of custom pattern families for A_{ex} as described in Plot Set Up 3.2. **Left:** $\mathcal{F}^1$ with multilinear patterns $ML(\alpha, \{0,1\}^2), \alpha \in \{(1,1),(5,5)\}$, the chains $CH(\alpha, 5)$, $\alpha \in \{\mathbf{e}^1, \mathbf{e}^2, \mathbf{1}^2\}$ and the shifted chains $(0,5) + CH(\mathbf{e}^1, 5)$ and $(2,0) + CH(\mathbf{e}^2, 5)$. **Middle:** $\mathcal{F}^2$ with multilinear patterns $ML(\alpha, \{0,1\}^2)$, $\alpha \in \{(1,1),(2,3),(2,4),(5,5)\}$, chains $CH(\alpha, 5)$, $\alpha \in \{\mathbf{e}^1, \mathbf{e}^2\}$. **Right:** $\mathcal{F}^3$ with multilinear patterns $ML(\alpha, \{0,1\}^2)$, $\alpha \in \{(1,1),(2,3),(2,4),(5,5)\}$, chains $CH(\alpha, 5), \alpha \in \{\mathbf{e}^1, \mathbf{e}^2, \mathbf{1}^2\}$ and shifted chains $(0,5) + CH(\mathbf{e}^1, 5)$, $(2,0) + CH(\mathbf{e}^2, 5)$ and $(0,4) + CH((2,0), 2)$.

While the bounds obtained from $\mathcal{F}^2$, see P^2 in Figure 6.7, are far from optimal, they are an improvement compared to $\mathcal{F}^m_{A_{ex}}$ in column (M). The more involved pattern families $\mathcal{F}^1$ and $\mathcal{F}^3$ result in similar bounds as (B) and (CS).

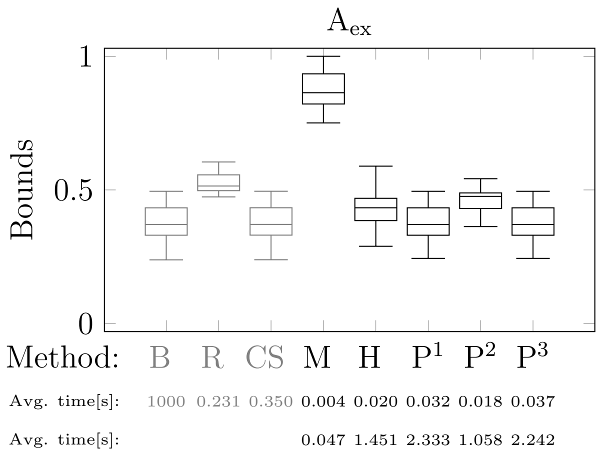

Figure 6.7: Visualization of the distributions of the normalized width functions (6.2) for various methods from Tables 6.1 and 6.2 as described in Box Plot Set Up 6.1 The custom pattern families $\mathcal{F}^1, \mathcal{F}^2, \mathcal{F}^3$ for A_{ex} labeled as P^1, P^2, P^3 are defined in Figure 6.6.

Chapter 7

Conclusion

We have presented a customizable framework for the relaxation of polynomial optimization problems that is based on monomial patterns. This novel framework allows to integrate approaches from different communities or to develop new approaches. In fact, we have shown that various methods using linearizations by multilinear terms or bound-factor products as well as relaxations based on SOS, SDSOS and SONC polynomials all come with their particular types of patterns. The advantage of our approach is that, by using patterns, we gain flexibility in terms of the size of the relaxation. By exploiting the combinatorial structure of the set A of monomial exponents we are able to neglect dependencies between certain monomials. This in turn allows to avoid hard problem formulations and instead focus on well-behaved and easy-to-describe dependencies between certain other monomials. The key idea is to replace the cone $\mathcal{P}(\mathrm{K})_{\mathrm{A}}$ by

$$\sum_{\mathrm{P}\in\mathcal{F}} \mathcal{P}(\mathrm{K})_{\mathrm{P}}, \tag{7.1}$$

where $\mathcal{F}$ is a family of patterns. By carefully choosing $\mathcal{F}$ we can produce tractable and sufficiently tight relaxations of (POP). While not explicitly expressed this way, SDSOS or SONC relaxations of (POP) successfully use this concept. Rigorously exploiting this idea we introduce truncated submonoids – a novel pattern type. Truncated submonoids $\mathrm{TS}(\eta, \Gamma, \mathrm{B})$ allow to control the cost of enforcing the constraint

$$f^{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} \in \mathcal{P}(\mathrm{K})_{\mathrm{TS}(\eta,\Gamma,\mathrm{B})} \tag{7.2}$$

by choosing the parameters Γ and B appropriately. From truncated submonoids we derive a variety of other pattern types such as chains. Chains pose a generalization of binomial square patterns, which is the underlying pattern of SDSOS relaxations. Recall that a binomial square pattern P links 3 monomial variables and that, when $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$, the constraint $f^{\mathrm{P}} \in \mathcal{P}(\mathrm{K})_{\mathrm{P}}$ can be enforced by PSD constraints involving one 2×2 matrix. Our generalization preserves the desirable properties of binomial square patterns: a chain $\mathrm{CH}(\eta, \gamma, 2\mathrm{d})$ allows to link $2\mathrm{d}+1$ monomial variables by

$$f^{\mathrm{CH}(\eta,\gamma,2\mathrm{d})} \in \mathcal{P}(\mathrm{K})_{\mathrm{CH}(\eta,\gamma,2\mathrm{d})}. \tag{7.3}$$

If $\mathrm{K} = \mathbb{R}^{\mathrm{n}}$, (7.3) can be enforced using one PSD matrix of size $\mathrm{d}+1 \times \mathrm{d}+1$. Thus, they induce constraints of manageable size, but also allow to link more monomial variables. Furthermore, if K is an axis-parallel box, (7.3) can be enforced by at most two PSD matrices of size at most $\mathrm{d}+1\times\mathrm{d}+1$. Hence, effectively creating an analogue of SDSOS cones for axis-parallel boxes K.

However, it is the author's opinion that an exact description of the cone in constraint (7.2) is not necessary. In fact, SOS, SDSOS and SONC relaxations or multilinear envelopes usually yield only inner approximations of $\mathcal{P}(\mathrm{K})_{\mathrm{A}}$. It therefore seems arbitrary to demand an exact description of $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ in (7.1). Dropping this demand allows to combine truncated submonoids with a variety of Positivstellensätze to create tractable relaxations of $\mathcal{P}(\mathrm{K})_{\mathrm{P}}$ and consequently of (POP). This allows to use the different established SOS/SDP-based approaches which in practice become quickly infeasible due to the involved computational cost of solving large SDPs.

Another merit of the pattern relaxation is that different pattern types can be used to compose a tractable relaxation of (POP). In fact, all discussed patterns in this thesis can be combined. This includes combinations of SONC and SOS, due to the chosen power cone representations of circuit polynomials. By combining different pattern types, we can avoid ill-behaved problem formulations.

Furthermore, once clearly articulated, certain basic and useful facts become very apparent within the pattern relaxation. By introducing the notions of 'directly connected' and 'indirectly connected' moment variables, we can express the problem of finding a well balanced pattern family $\mathcal{F}$ as finding a minimal node-weighted Steiner tree within a certain graph. In Example 5.6 the graph approach is applied in order to find a family of multilinear patterns.

The computational studies clearly provide numerical evidence that it is worth to think about an appropriate choice of pattern types and their combination. The different families show how their compositions influence the quality of the lower bounds. While we are able to generate reasonably tight lower bounds from pattern relaxations when suitable pattern families are used, the opposite holds true for ill-fitted pattern families. We want to point out that the way we choose particular pattern families $\mathcal{F}$ and handle these patterns computationally in the numerical evaluation is merely one of the many possible options.

In particular, the cutting-plane algorithm 5.1 allows to combine the different approaches in numerous ways. The cuts generated by (SP) can be integrated directly into branch-and-bound that use monomial variables such as `BARON` [66], `SCIP` [24], `COUENNE` [12] or `LINDOGlobal` [57]. Thus, by choosing an appropriate set of generators of a truncated submonoid, (SP) can be used as an interface for SOS methods in branch-and-bound algorithms.

7.1 Outlook

There are different ways to improve the pattern relaxation. While this thesis shows that there already exists a rather large toolbox of patterns, expanding it will almost certainly stimulate the creation of new relaxations.

Another pressing issue is how to determine an appropriate pattern family $\mathcal{F}$ for a given problem. This could be achieved in the following ways:

- The development of graph based approaches as in Example 5.6 to generate customized families for each problem.
- The development of iterative procedures, that allow to augment a given pattern family after the optimal solution of the corresponding pattern relaxation is computed.
- Identifying combinations of patterns that complement each other. An example of this is the combination of chains and multilinear patterns in $\mathcal{F}_{\mathrm{A}}^{\mathrm{h}}$, see Section 6.2.

In order to solve challenging polynomial problems at the frontier of what is currently possible, it seems to be necessary to merge the approaches of different communities. A promising prospect for this is the combination of branch-and-bound methods with sophisticated relaxation methods from the polynomial optimization community. The relaxations that are usually solved within branch-and-bound methods in order to compute lower bounds, come along with relatively low computational cost. This allows to compute many bounds, which helps to mitigate the effects of their (usually) low quality. However, if the quality of these bounds is too low, the branch-and-bound approach is not feasible anymore. Here, pattern relaxations can be used to fit the tightness and the computational cost of the lower bounds to the demands of branch-and-bound algorithms. Figure 7.1 presents how a spatial branch-and-bound algorithm can be enhanced using the pattern relaxation (D-RLX) to compute lower bounds. Additionally, the augmentation step (Augment relaxation by adding patterns to $\mathcal{F}$) in Figure 7.1 yields the option to strengthen the relaxation if a lower bound is not sufficiently tight for further branching. However, at this point it is not clear how to determine whether the bounds are sufficiently tight nor which pattern should be added to $\mathcal{F}$ in order to generate tighter bounds.

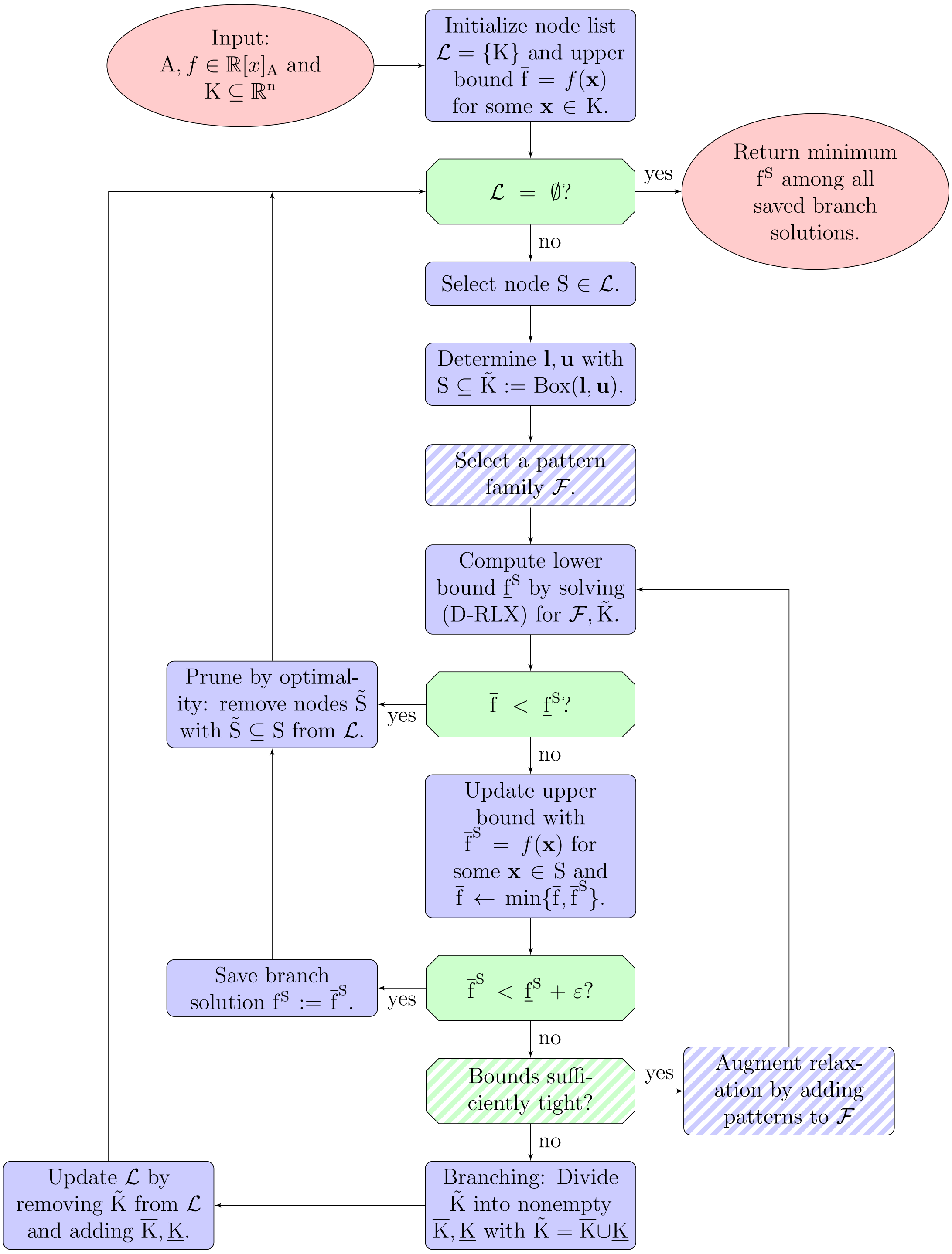

Figure 7.1: Flowchart (adapted from [32, Fig. 2]) of a prototype branch-and-bound procedure enhanced by the pattern relaxation approach for solving (POP) with objective function $f \in \mathbb{R}[x]_A$ and a compact feasible set K. The boxes that are striped showcase open problems. In particular how to decide whether the bounds are sufficiently tight for further branching and how to select new patterns to augment $\mathcal{F}$ are open questions.

Danksagung

Nomenclature and Notation

The following list of symbols complements the notation introduced in Section 1.2. Let n be a positive and d a nonnegative integer. Furthermore, let $\mathrm{A} \subseteq \mathbb{N}^{\mathrm{n}}$ be a nonempty and finite set, $\mathcal{F}$ be a pattern family, $\mathrm{K} \subseteq \mathbb{R}^{\mathrm{n}}$ be a closed set that contains a full-dimensional ball $\{\mathbf{x} \in \mathbb{R}^{\mathrm{n}} : ||\mathbf{x} - \mathbf{a}||_1 < \varepsilon\}$ for some $\varepsilon > 0$ and $\mathbf{a} \in \mathbb{R}^{\mathrm{n}}$, $\mathrm{X} \subseteq \mathbb{R}^{\mathrm{A}}$ be a nonempty and compact set, f be a polynomial and $\mathbf{w} \in \mathbb{R}^{\mathrm{A}}$ be a vector.

List of Symbols

$\triangle$	end of a definition, lemma, or theorem
$\square$	end of a proof
$\mathbb{N}$	set of natural numbers including zero
$\mathbb{N}^{\mathrm{n}}_{\mathrm{d}}$	$:= \{\alpha \in \mathbb{N}^{\mathrm{n}} : \alpha_1 + \cdots + \alpha_{\mathrm{n}} \leq \mathrm{d}\}$
$\mathbb{N}^{\mathrm{n}}_{\beta}$	$:= \{\alpha \in \mathbb{N}^{\mathrm{n}} : \alpha_{\mathrm{i}} \leq \beta_{\mathrm{i}}$ for all $\mathrm{i} \in [\mathrm{n}]\}$ for $\beta \in \mathbb{N}^{\mathrm{n}}$
$\mathbb{P}_{\mu}$	power cone, page 40
$\mathbb{R}$	set of real numbers
$\mathbb{R}_{+}$	$:= \{\mathrm{v} \in \mathbb{R} : \mathrm{v} \geq 0\}$
$\mathbb{R}_{++}$	$:= \{\mathrm{v} \in \mathbb{R} : \mathrm{v} > 0\}$
$\mathbb{R}^{\mathrm{A}}$	$:= \{(\mathrm{v}_{\alpha})_{\alpha \in \mathrm{A}} : \mathrm{v}_{\alpha} \in \mathbb{R}$ for all $\alpha \in \mathrm{A}\} \cong \mathbb{R}^{\#\mathrm{A}}$
$\mathbb{R}[x]$	$:= \{\sum_{\alpha \in \mathbb{N}^{\mathrm{n}}} \mathrm{f}_{\alpha} x^{\alpha} : \mathbf{f} \in \mathbb{R}^{\mathbb{N}^{\mathrm{n}}}$ with $\#\,\mathrm{supp}(\mathbf{f}) < \infty\}$
$\mathbb{R}[x]_{\mathrm{A}}$	$:= \{\sum_{\alpha \in \mathrm{A}} \mathrm{f}_{\alpha} x^{\alpha} : \mathbf{f} \in \mathbb{R}^{\mathrm{A}}\}$
$\mathbb{S}^{\mathrm{A}}_{+}$	$:= \{\mathbf{M} \in \mathbb{R}^{\mathrm{A} \times \mathrm{A}} : \mathbf{M}$ is symmetric and positive semidefinite$\}$
$\mathbb{Z}$	set of integers
$\mathcal{C}(\mathrm{K})_{\mathrm{A}}$	$\mathrm{cl\,cone}(\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\})$ is the A-truncated moment cone
$\mathcal{M}(\mathrm{K})_{\mathrm{A}}$	$\mathrm{cl\,conv}(\{\mathrm{m}(\mathbf{x})_{\mathrm{A}} : \mathbf{x} \in \mathrm{K}\})$ is the A-truncated moment body
$\mathcal{P}(\mathrm{K})$	$:= \{f \in \mathbb{R}[x] : f \geq 0$ on $\mathrm{K}\}$

$\mathcal{P}(\mathrm{K})_\mathrm{A}$	$:= \{f \in \mathcal{P}(\mathrm{K}) : \mathrm{supp}(f) \subseteq \mathrm{A}\}$
$[\mathrm{n}]$	$:= \{1, 2, \ldots, \mathrm{n}\}$
$[\mathrm{n}]^\mathrm{m}$	$:= \bigtimes_{\mathrm{i}\in[\mathrm{m}]}[\mathrm{n}]$ for $\mathrm{m} \in \mathbb{N}\setminus\{0\}$
$[\mathrm{n}]_0$	$:= \{0, 1, 2, \ldots, \mathrm{n}\}$
$[\mathrm{n}]_0^\mathrm{m}$	$:= \bigtimes_{\mathrm{i}\in[\mathrm{m}]}[\mathrm{n}]_0$ for $\mathrm{m} \in \mathbb{N}\setminus\{0\}$
$\{\mathrm{a}_\beta\}_{\beta\in\mathrm{B}}$	$:= \{\mathrm{a}_\beta : \beta \in \mathrm{B}\}$ for $\mathrm{B} \subseteq \mathbb{N}^\mathrm{n}$
$\#$	cardinality
$\circ$	Hadamard product, page 25
$\dot\cup$	disjoint union
$\cong$	isomorph
$\leq, \geq$	interpreted componentwise for vectors
$<, >$	interpreted componentwise for vectors
$\lvert\cdot\rvert$	absolute value
$\lVert\cdot\rVert_1$	ℓ_1-norm
$\lVert f\rVert_1$	$:= \lVert \mathrm{vec}(f)_{\mathrm{supp}(f)}\rVert_1$
$\lVert\cdot\rVert_\infty$	ℓ_∞-norm
$\emptyset$	empty set
$\varrho(\mathrm{I})$	fineness of I
$\Phi^{[\mathrm{a},\mathrm{b}]}$	page 72
$[\![\mathrm{X}]\!]_\mathrm{B}$	$:= \{\mathbf{v}_\mathrm{B} : \mathbf{v} \in \mathrm{X}\}$ for nonempty $\mathrm{B} \subseteq \mathrm{A}$
$\{0,1\}^\mathrm{A}$	$:= \bigtimes_{\alpha\in\mathrm{A}} \{0,1\}$
$[\mathrm{a},\mathrm{b}]^\mathrm{A}$	$:= \bigtimes_{\alpha\in\mathrm{A}} [\mathrm{a},\mathrm{b}]$ for $\mathrm{a},\mathrm{b} \in \mathbb{R}$
$\mathbf{e}^{\alpha,\mathrm{A}}$ or $\mathbf{e}^\alpha$	standard basis vectors of $\mathbb{R}^\mathrm{A}$ for $\alpha \in \mathrm{A}$
$\mathbf{1}^\mathrm{A}$ or $\mathbf{1}$	all ones vector in $\mathbb{R}^\mathrm{A}$
$\mathbf{0}^\mathrm{A}$ or $\mathbf{0}$	all zeros vector in $\mathbb{R}^\mathrm{A}$
$\mathbf{v}_\mathrm{B}$	$:= (\mathrm{v}_\alpha)_{\alpha\in\mathrm{B}}$ for $\mathrm{B} \subseteq \mathrm{A}, \mathbf{v} \in \mathbb{R}^\mathrm{A}$
$\mathrm{A}_\mathcal{F}$	$:= \bigcup_{\mathrm{P}\in\mathcal{F}} \mathrm{P}$

$\mathrm{Box}(\mathbf{l},\mathbf{u})$	$:= [\mathrm{l}_1,\mathrm{u}_1] \times \cdots \times [\mathrm{l}_\mathrm{n},\mathrm{u}_\mathrm{n}]$ for $\mathbf{l},\mathbf{u} \in \mathbb{R}^\mathrm{n}$ with $\mathbf{l} \leq \mathbf{u}$
cl	closure
cone	conic hull
conv	convex hull
co(P)	computational cost of a pattern P, page 88
$\deg(\alpha)$	$:= \alpha_1 + \cdots + \alpha_\mathrm{n}$ is the degree of $\alpha \in \mathbb{N}^\mathrm{n}$
$\deg(\mathrm{A})$	$:= \max\{\deg(\beta) : \beta \in \mathrm{A}\}$ is the degree of A
$\deg(f)$	$:= \deg(\mathrm{supp}(f))$ degree of $f \in \mathbb{R}[x]$
$\det(\mathbf{T})$	determinat of a matrix $\mathbf{T} \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$
$\mathrm{diag}(\mathbf{w})$	$\mathrm{A} \times \mathrm{A}$ diagonal matrix with diagonal $\mathbf{w}$
$\mathrm{diam}(\mathrm{X})$	$:= \max\limits_{\mathbf{u},\mathbf{z}\in\mathrm{X}} \lVert\mathbf{u}-\mathbf{z}\rVert_1$
$\mathrm{dist}(\mathrm{X},\mathbf{w})$	$:= \min\limits_{\mathbf{u}\in\mathrm{X}} \lVert\mathbf{w}-\mathbf{u}\rVert_1$
Ed (G)	edges of the graph G, page 89
$\mathrm{G}(\mathcal{A})$	graph induces by the pattern family $\mathcal{A}$, page 89
$\mathrm{m}(x)_\mathrm{A}$	$:= (x^\alpha)_{\alpha\in\mathrm{A}}$ is the A-truncated moment vector map
$\mathrm{N}_\varepsilon(\mathrm{X})$	$:= \{\mathbf{u} \in \mathbb{R}^\mathrm{A} : \lVert\mathbf{u}-\mathbf{z}\rVert_1 \leq \varepsilon$ for some $\mathbf{z} \in \mathrm{X}\}$ for $\varepsilon > 0$
No (G)	nodes of the graph G, page 89
relint	relative interior
SOS(A)	$:= \{p \in \mathbb{R}[x]_\mathrm{A} : p$ is SOS$\}$
$\mathrm{supp}(\mathbf{w})$	$:= \{\alpha \in \mathrm{A} : \mathrm{w}_\alpha \neq 0\}$ support of $\mathbf{w}$
$\mathrm{supp}(\mathrm{X})$	$:= \bigcup_{\mathbf{v}\in\mathrm{X}} \mathrm{supp}(\mathbf{v})$ support of X
$\mathrm{supp}(f)$	$:= \mathrm{supp}(\mathrm{vec}(f)_{\mathbb{N}^\mathrm{n}})$ monomial support of f
$\mathrm{vec}(f)_\mathrm{A}$	$:= \mathbf{f}_\mathrm{A}$ coefficients of $f = \sum_{\alpha\in\mathbb{N}^\mathrm{n}} \mathrm{f}_\alpha x^\alpha$ indexed by A
$\mathrm{vec}(f)_\alpha$	$:= \mathrm{vec}(f)_{\{\alpha\}}$
we(a)	weight of node $\mathrm{a} \in \mathrm{G}(\mathcal{A})$, page 89
$\mathrm{width}(\mathrm{X},\mathbf{c})$	$:= \max\{\langle\mathbf{c},\mathbf{u}\rangle : \mathbf{u} \in \mathrm{X}\} - \min\{\langle\mathbf{c},\mathbf{u}\rangle : \mathbf{u} \in \mathrm{X}\}$ for $\mathbf{c} \in \mathbb{R}^\mathrm{A}$
x^α	$:= x_1^{\alpha_1} \cdot \ldots \cdot x_\mathrm{n}^{\alpha_\mathrm{n}}$ for an indeterminate $x = (x_1,\ldots,x_\mathrm{n})$

$\mathbf{x}^{\alpha}$	$:= x_1^{\alpha_1} \cdot \ldots \cdot x_n^{\alpha_n}$ for $\mathbf{x} \in \mathbb{R}^n$
$\overline{\mathbf{x}}_K^{\alpha}$	$:= \max_{\mathbf{x} \in K} \mathbf{x}^{\alpha}$
$\overline{\mathbf{x}}_K^{A}$	$:= (\overline{\mathbf{x}}_K^{\alpha})_{\alpha \in A}$
$\underline{\mathbf{x}}_K^{\alpha}$	$:= \min_{\mathbf{x} \in K} \mathbf{x}^{\alpha}$
$\underline{\mathbf{x}}_K^{A}$	$:= (\underline{\mathbf{x}}_K^{\alpha})_{\alpha \in A}$

List of Acronyms

List of Figures

List of Tables

Bibliography

[1] A. A. Ahmadi and A. Majumdar. DSOS and SDSOS optimization: more tractable alternatives to sum of squares and semidefinite optimization. *SIAM J. Appl. Algebra Geom.*, 3(2):193–230, 2019. doi: 10.1137/18M118935X. URL `https://doi.org/10.1137/18M118935X`.

[2] M. F. Anjos and J. B. Lasserre, editors. *Handbook on semidefinite, conic and polynomial optimization*, volume 166 of *International Series in Operations Research & Management Science*. Springer, New York, 2012. ISBN 978-1-4614-0768-3. doi: 10.1007/978-1-4614-0769-0. URL `https://doi.org/10.1007/978-1-4614-0769-0`.

[3] M. ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 9.0.*, 2019. URL `http://docs.mosek.com/9.0/toolbox/index.html`.

[4] M. ApS. *MOSEK Modeling Cookbook*, 2020. URL `https://docs.mosek.com/modeling-cookbook/index.html`.

[5] E. Artin. Über die Zerlegung definiter Funktionen in Quadrate. *Abh. Math. Sem. Univ. Hamburg*, 5(1):100–115, 1927. ISSN 0025-5858. doi: 10.1007/BF02952513. URL `https://doi.org/10.1007/BF02952513`.

[6] G. Averkov. Constructive proofs of some Positivstellensätze for compact semialgebraic subsets of $\mathbb{R}^d$. *J. Optim. Theory Appl.*, 158(2):410–418, 2013. ISSN 0022-3239. doi: 10.1007/s10957-012-0261-9. URL `https://doi.org/10.1007/s10957-012-0261-9`.

[7] G. Averkov. Optimal size of linear matrix inequalities in semidefinite approaches to polynomial optimization. *SIAM J. Appl. Algebra Geom.*, 3(1):128–151, 2019. doi: 10.1137/18M1201342. URL `https://doi.org/10.1137/18M1201342`.

[8] G. Averkov. Second-order cone representable slices of the positive semidefinite cone of size three. *arXiv preprint arXiv:1909.08937*, 2019.

[9] G. Averkov, B. Peters, and S. Sager. Convexification of polynomial optimization problems by means of monomial patterns. *arXiv preprint arXiv:1901.05675*, 2019.

[10] T. Bajbar. *On Newton polytopes and growth properties of multivariate polynomials*. PhD thesis, Karlsruher Institut für Technologie, 7 2016.

[11] X. Bao, A. Khajavirad, N. V. Sahinidis, and M. Tawarmalani. Global optimization of nonconvex problems with multilinear intermediates. *Math. Program. Comput.*, 7(1):1–37, 2015. ISSN 1867-2949. doi: 10.1007/s12532-014-0073-z. URL `https://doi.org/10.1007/s12532-014-0073-z`.

[12] P. Belotti. Couenne, an exact solver for nonconvex minlps, 2015. URL `https://projects.coin-or.org/Couenne/`.

[13] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017. doi: 10.1137/141000671.

[14] G. Blekherman, P. A. Parrilo, and R. R. Thomas, editors. *Semidefinite optimization and convex algebraic geometry*, volume 13 of *MOS-SIAM Series on Optimization*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA; Mathematical Optimization Society, Philadelphia, PA, 2013. ISBN 978-1-611972-28-3.

[15] N. Boland, S. S. Dey, T. Kalinowski, M. Molinaro, and F. Rigterink. Bounding the gap between the mccormick relaxation and the convex hull for bilinear functions. *Mathematical Programming*, 162(1-2):523–535, 2017.

[16] S. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge University Press, Cambridge, 2004. ISBN 0-521-83378-7. doi: 10.1017/CBO9780511804441. URL `https://doi.org/10.1017/CBO9780511804441`.

[17] V. Chandrasekaran and P. Shah. Relative entropy relaxations for signomial optimization. *SIAM J. Optim.*, 26(2):1147–1173, 2016. ISSN 1052-6234. doi: 10.1137/140988978. URL `https://doi.org/10.1137/140988978`.

[18] R. Chares. *Cones and interior-point algorithms for structured convex optimization involving powers and exponentials*. PhD thesis, PhD thesis, Ph. D. Thesis, UCL-Université Catholique de Louvain, Louvain-la-Neuve, Belgium, 2009.

[19] E. Dalkiran and H. D. Sherali. Theoretical filtering of RLT bound-factor constraints for solving polynomial programming problems to global optimality. *J. Global Optim.*, 57(4):1147–1172, 2013. ISSN 0925-5001. doi: 10.1007/s10898-012-0024-z. URL `https://doi.org/10.1007/s10898-012-0024-z`.

[20] A. Del Pia and A. Khajavirad. A polyhedral study of binary polynomial programs. *Math. Oper. Res.*, 42(2):389–410, 2017. ISSN 0364-765X. doi: 10.1287/moor.2016.0804. URL `https://doi.org/10.1287/moor.2016.0804`.

[21] M. Dressler, S. Iliman, and T. de Wolff. A Positivstellensatz for sums of nonnegative circuit polynomials. *SIAM J. Appl. Algebra Geom.*, 1(1):536–555, 2017. ISSN 2470-6566. doi: 10.1137/16M1086303. URL `https://doi.org/10.1137/16M1086303`.

[22] M. Dressler, J. Heuer, H. Naumann, and T. de Wolff. Global optimization via the dual sonc cone and linear programming. *arXiv preprint arXiv:2002.09368*, 2020.

[23] G. Gamrath, T. Koch, S. J. Maher, D. Rehfeldt, and Y. Shinano. SCIP-Jack—a solver for STP and variants with parallelization extensions. *Math. Program. Comput.*, 9(2):231–296, 2017. ISSN 1867-2949. doi: 10.1007/s12532-016-0114-x. URL `https://doi.org/10.1007/s12532-016-0114-x`.

[24] G. Gamrath, D. Anderson, K. Bestuzheva, W.-K. Chen, L. Eifler, M. Gasse, P. Gemander, A. Gleixner, L. Gottwald, K. Halbig, G. Hendel, C. Hojny, T. Koch, P. Le Bodic, S. J. Maher, F. Matter, M. Miltenberger, E. Mühmer, B. Müller, M. E. Pfetsch, F. Schlösser, F. Serrano, Y. Shinano, C. Tawfik, S. Vigerske, F. Wegscheider, D. Weninger, and J. Witzig. The SCIP Optimization Suite 7.0. ZIB-Report 20-10, Zuse Institute Berlin, March 2020. URL `http://nbn-resolving.de/urn:nbn:de:0297-zib-78023`.

[25] B. Gärtner and J. Matoušek. *Approximation algorithms and semidefinite programming*. Springer, Heidelberg, 2012. ISBN 978-3-642-22014-2; 978-3-642-22015-9. doi: 10.1007/978-3-642-22015-9. URL `https://doi.org/10.1007/978-3-642-22015-9`.

[26] F. Glover and E. Woolsey. Converting the 0-1 polynomial programming problem to a 0-1 linear program. *Operations research*, 22(1):180–182, 1974.

[27] D. Handelman. Representing polynomials by positive linear functions on compact convex polyhedra. *Pacific J. Math.*, 132(1):35–62, 1988. ISSN 0030-8730. URL `http://projecteuclid.org/euclid.pjm/1102689794`.

[28] C. Helmberg and F. Rendl. A spectral bundle method for semidefinite programming. *SIAM J. Optim.*, 10(3):673–696, 2000. ISSN 1052-6234. doi: 10.1137/S1052623497328987. URL `https://doi.org/10.1137/S1052623497328987`.

[29] S. Iliman and T. de Wolff. Amoebas, nonnegative polynomials and sums of squares supported on circuits. *Res. Math. Sci.*, 3:Paper No. 9, 35, 2016. ISSN 2522-0144. doi: 10.1186/s40687-016-0052-2. URL `https://doi.org/10.1186/s40687-016-0052-2`.

[30] C. Josz and D. K. Molzahn. Lasserre hierarchy for large scale polynomial optimization in real and complex variables. *SIAM J. Optim.*, 28(2):1017–1048, 2018. ISSN 1052-6234. doi: 10.1137/15M1034386. URL https://doi.org/10.1137/15M1034386.

[31] L. Katthän, H. Naumann, and T. Theobald. A unified framework of sage and sonc polynomials and its duality theory. *arXiv preprint arXiv:1903.08966*, 2019.

[32] K. Kianfar. Branch-and-bound algorithms. *Wiley Encyclopedia of Operations Research and Management Science*, 2010.

[33] B. Korte and J. Vygen. *Combinatorial optimization*, volume 21 of *Algorithms and Combinatorics*. Springer, Berlin, 2018. ISBN 978-3-662-56038-9; 978-3-662-56039-6. doi: 10.1007/978-3-662-56039-6. URL https://doi.org/10.1007/978-3-662-56039-6. Theory and algorithms, Sixth edition of [MR1764207].

[34] J.-L. Krivine. Anneaux préordonnés. *J. Analyse Math.*, 12:307–326, 1964. ISSN 0021-7670. doi: 10.1007/BF02807438. URL https://doi.org/10.1007/BF02807438.

[35] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM J. Optim.*, 11(3):796–817, 2000/01. ISSN 1052-6234. doi: 10.1137/S1052623400366802. URL https://doi.org/10.1137/S1052623400366802.

[36] J. B. Lasserre. Convergent SDP-relaxations in polynomial optimization with sparsity. *SIAM J. Optim.*, 17(3):822–843, 2006. ISSN 1052-6234. doi: 10.1137/05064504X. URL https://doi.org/10.1137/05064504X.

[37] J. B. Lasserre. *An introduction to polynomial and semi-algebraic optimization.* Cambridge Texts in Applied Mathematics. Cambridge University Press, Cambridge, 2015. ISBN 978-1-107-63069-7; 978-1-107-06057-9. doi: 10.1017/CBO9781107447226. URL https://doi.org/10.1017/CBO9781107447226.

[38] M. Laurent. Sums of squares, moment matrices and optimization over polynomials. In *Emerging applications of algebraic geometry*, volume 149 of *IMA Vol. Math. Appl.*, pages 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.

[39] V. Magron, H. Seidler, and T. de Wolff. Exact optimization via sums of nonnegative circuits and arithmetic-geometric-mean-exponentials. In *ISSAC'19—Proceedings of the 2019 ACM International Symposium on Symbolic and Algebraic Computation*, pages 291–298. ACM, New York, 2019. doi: 10.1145/3326229.3326271. URL https://doi.org/10.1145/3326229.3326271.

[40] M. Marshall. *Positive polynomials and sums of squares*, volume 146 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2008. ISBN 978-0-8218-4402-1; 0-8218-4402-4. doi: 10.1090/surv/146. URL `https://doi.org/10.1090/surv/146`.

[41] *MATLAB version 9.6.0.1174912 (R2019a) Update 5*. The Mathworks, Inc., Natick, Massachusetts, 2019.

[42] G. P. McCormick. Computability of global solutions to factorable nonconvex programs. I. Convex underestimating problems. *Math. Programming*, 10(2):147–175, 1976. ISSN 0025-5610. doi: 10.1007/BF01580665. URL `https://doi.org/10.1007/BF01580665`.

[43] A. Mitsos, B. Chachuat, and P. Barton. 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`.

[44] K. G. Murty and S. N. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Math. Programming*, 39(2):117–129, 1987. ISSN 0025-5610. doi: 10.1007/BF02592948. URL `https://doi.org/10.1007/BF02592948`.

[45] Y. Nesterov. Squared functional systems and optimization problems. In *High performance optimization*, volume 33 of *Appl. Optim.*, pages 405–440. Kluwer Acad. Publ., Dordrecht, 2000. doi: 10.1007/978-1-4757-3216-0_17. URL `https://doi.org/10.1007/978-1-4757-3216-0_17`.

[46] J. Nie. Optimality conditions and finite convergence of Lasserre's hierarchy. *Math. Program.*, 146(1-2, Ser. A):97–121, 2014. ISSN 0025-5610. doi: 10.1007/s10107-013-0680-x. URL `https://doi.org/10.1007/s10107-013-0680-x`.

[47] D. Papp. Duality of sum of nonnegative circuit polynomials and optimal SONC bounds. *arXiv preprint arXiv:1912.04718*, 2019.

[48] P. A. Parrilo. Semidefinite programming relaxations for semialgebraic problems. volume 96, pages 293–320. 2003. doi: 10.1007/s10107-003-0387-5. URL `https://doi.org/10.1007/s10107-003-0387-5`. Algebraic and geometric methods in discrete optimization.

[49] P. A. Parrilo and B. Sturmfels. Minimizing polynomial functions. In *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, volume 60 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 83–99. Amer. Math. Soc., Providence, RI, 2003.

[50] Y. Puranik and N. V. Sahinidis. Domain reduction techniques for global NLP and MINLP optimization. *Constraints*, 22(3):338–376, 2017. ISSN 1383-7133. doi: 10.1007/s10601-016-9267-5. URL `https://doi.org/10.1007/s10601-016-9267-5`.

[51] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana Univ. Math. J.*, 42(3):969–984, 1993. ISSN 0022-2518. doi: 10.1512/iumj.1993.42.42045. URL `https://doi.org/10.1512/iumj.1993.42.42045`.

[52] D. Rehfeldt, T. Koch, and S. J. Maher. Reduction techniques for the prize collecting Steiner tree problem and the maximum-weight connected subgraph problem. *Networks*, 73(2):206–233, 2019. ISSN 0028-3045. doi: 10.1002/net.21857. URL `https://doi.org/10.1002/net.21857`.

[53] C. Riener, T. Theobald, L. J. Andrén, and J. B. Lasserre. Exploiting symmetries in SDP-relaxations for polynomial optimization. *Math. Oper. Res.*, 38(1):122–141, 2013. ISSN 0364-765X. doi: 10.1287/moor.1120.0558. URL `https://doi.org/10.1287/moor.1120.0558`.

[54] R. T. Rockafellar. *Convex analysis*. Princeton Mathematical Series, No. 28. Princeton University Press, Princeton, N.J., 1970.

[55] N. V. Sahinidis. *BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs,* User's Manual, 2017.

[56] R. Schneider. *Convex bodies: the Brunn-Minkowski theory*, volume 151 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, expanded edition, 2014. ISBN 978-1-107-60101-7.

[57] L. E. Schrage. *Optimization modeling with LINGO*. Lindo System, 2006.

[58] H. Seidler and T. de Wolff. An experimental comparison of sonc and sos certificates for unconstrained optimization. *arXiv preprint arXiv:1808.08431*, 2018.

[59] H. D. Sherali and W. P. Adams. *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, volume 31 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, 1999. ISBN 0-7923-5487-7. doi: 10.1007/978-1-4757-4388-3. URL `https://doi.org/10.1007/978-1-4757-4388-3`.

[60] N. Z. Shor. A class of estimates for the global minimum of polynomial functions. *Kibernetika (Kiev)*, (6):9–11, 133, 1987. ISSN 0023-1274.

[61] T. S. Shores. *Applied linear algebra and matrix analysis*. Undergraduate Texts in Mathematics. Springer, New York, 2007. ISBN 978-0-387-33194-2; 0-387-33194-8. doi: 10.1007/978-0-387-48947-6. URL https://doi.org/10.1007/978-0-387-48947-6.

[62] E. Smith, C. Pantelides, and G. Reklaitis. 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.

[63] E. Speakman and J. Lee. Quantifying double McCormick. *Math. Oper. Res.*, 42(4):1230–1253, 2017. ISSN 0364-765X. doi: 10.1287/moor.2017.0846. URL https://doi.org/10.1287/moor.2017.0846.

[64] G. Stengle. A nullstellensatz and a positivstellensatz in semialgebraic geometry. *Math. Ann.*, 207:87–97, 1974. ISSN 0025-5831. doi: 10.1007/BF01362149. URL https://doi.org/10.1007/BF01362149.

[65] O. Tange. GNU parallel 20201022 ('samuelpaty'), Oct. 2020. URL https://doi.org/10.5281/zenodo.4118697. GNU Parallel is a general parallelizer to run multiple serial command line programs in parallel without changing them.

[66] M. Tawarmalani and N. V. Sahinidis. A polyhedral branch-and-cut approach to global optimization. *Math. Program.*, 103(2, Ser. B):225–249, 2005. ISSN 0025-5610. doi: 10.1007/s10107-005-0581-8. URL https://doi.org/10.1007/s10107-005-0581-8.

[67] J. Wang, V. Magron, and J.-B. Lasserre. Chordal-TSSOS: A Moment-SOS Hierarchy That Exploits Term Sparsity with Chordal Extension. *SIAM J. Optim.*, 31(1):114–141, 2021. ISSN 1052-6234. doi: 10.1137/20M1323564. URL https://doi.org/10.1137/20M1323564.

[68] J. Wang, V. Magron, and J.-B. Lasserre. TSSOS: A Moment-SOS Hierarchy That Exploits Term Sparsity. *SIAM J. Optim.*, 31(1):30–58, 2021. ISSN 1052-6234. doi: 10.1137/19M1307871. URL https://doi.org/10.1137/19M1307871.

[69] T. Weisser, J. B. Lasserre, and K.-C. Toh. Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity. *Math. Program. Comput.*, 10(1):1–32, 2018. ISSN 1867-2949. doi: 10.1007/s12532-017-0121-6. URL https://doi.org/10.1007/s12532-017-0121-6.