Applications of convex analysis to signomial and polynomial nonnegativity problems

Thesis by
Riley John Murray

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2021
Defended May 13, 2021

ACKNOWLEDGEMENTS

iv

ABSTRACT

Here is a question that is easy to state, but often hard to answer:

Is this function nonnegative on this set?

When faced with such a question, one often makes appeals to known inequalities. One crafts arguments that are *sufficient* to establish the nonnegativity of the function, rather than determining the function's precise range of values. This thesis studies sufficient conditions for nonnegativity of signomials and polynomials. Conceptually, signomials may be viewed as generalized polynomials that feature arbitrary real exponents, but with variables restricted to the positive orthant.

Our methods leverage efficient algorithms for a type of convex optimization known as relative entropy programming (REP). By virtue of this integration with REP, our methods can help answer questions like the following:

Is there some function, in this particular space of functions, that is nonnegative on this set?

The ability to answer such questions is *extremely* useful in applied mathematics. Alternative approaches in this same vein (e.g., methods for polynomials based on semidefinite programming) have been used successfully as convex relaxation frameworks for nonconvex optimization, as mechanisms for analyzing dynamical systems, and even as tools for solving nonlinear partial differential equations.

This thesis builds from the *sums of arithmetic-geometric exponentials* or *SAGE* approach to signomial nonnegativity. The term "exponential" appears in the SAGE acronym because SAGE parameterizes signomials in terms of exponential functions.

Our first round of contributions concern the original SAGE approach. We employ basic techniques in convex analysis and convex geometry to derive structural results for spaces of SAGE signomials and exactness results for SAGE-based REP relaxations of nonconvex signomial optimization problems. We frame our analysis primarily in terms of the coefficients of a signomial's basis expansion rather than in terms of signomials themselves. The effect of this framing is that our results for signomials readily transfer to polynomials. In particular, we are led to define a new concept of *SAGE polynomials*. For sparse polynomials, this method offers an

exponential efficiency improvement relative to certificates of nonnegativity obtained through semidefinite programming.

We go on to create the *conditional SAGE* methodology for exploiting convex substructure in constrained signomial nonnegativity problems. The basic insight here is that since the standard relative entropy representation of SAGE signomials is obtained by a suitable application of convex duality, we are free to add additional convex constraints into the duality argument. In the course of explaining this idea we provide some illustrative examples in signomial optimization and analysis of chemical dynamics.

The majority of this thesis is dedicated to exploring fundamental questions surrounding conditional SAGE signomials. We approach these questions through analysis frameworks of *sublinear circuits* and *signomial rings*. These sublinear circuits generalize simplicial circuits of affine-linear matroids, and lead to rich modes of analysis for sets that are simultaneously convex in the usual sense and convex under a logarithmic transformation. The concept of signomial rings lets us develop a powerful signomial Positivstellensatz and an elementary signomial moment theory. The Positivstellensatz provides for an effective hierarchy of REP relaxations for approaching the value of a nonconvex signomial minimization problem from below, as well as a first-of-its-kind hierarchy for approaching the same value from above.

In parallel with our mathematical work, we have developed the `sageopt` python package. `Sageopt` drives all the examples and experiments used throughout this thesis, and has been used by engineers to solve high-degree polynomial optimization problems at scales unattainable by alternative methods. We conclude this thesis with an explanation of how our theoretical results affected `sageopt`'s design.

PUBLISHED CONTENT AND CONTRIBUTIONS

Murray, R., Chandrasekaran, V., Wierman, A. (2021). "Newton polytopes and relative entropy optimization". In: *Foundations of Computational Mathematics*. DOI: `10.1007/s10208-021-09497-w`. arXiv pre-print (2018). R.M. participated in the conception of the project, performed the analysis, and participated in the writing of the manuscript.

Murray, R., Chandrasekaran, V., Wierman, A. (2020). "Signomial and polynomial optimization via relative entropy and partial dualization". In: *Mathematical Programming Computation*. DOI: `10.1007/s12532-020-00193-4`. arXiv pre-print (2019). R.M. participated in the conception of the project, performed the analysis, developed the software, undertook the experiments, and participated in the writing of the manuscript.

Murray, R., Naumann, H., Theobald, T. "Sublinear circuits and the constrained signomial nonnegativity problem." arXiv pre-print (2020). R.M. participated in the conception of the project, the analysis, and the writing of the manuscript.

Dressler, M., Murray, R. "An algebraic approach to signomial optimization." In preparation (2021). R.M. conceived the project, created the analysis framework, performed the analysis of the lower bounds, performed the experiments, participated in the analysis of the upper bounds, and participated in the writing of the manuscript.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

C h a p t e r   1

INTRODUCTION

This thesis considers intractable convex optimization problems. The construction of these problems begins by considering a linear space of real-valued functions, such as multivariate polynomials in some number of variables. We then consider the set of all functions in that space that are nonnegative on a domain of interest. These sets are called *nonnegativity cones*. They are evidently convex, because if we start with functions that are nonnegative on a domain K, then any positively weighted sum of those functions is also nonnegative on K.

One can get a sense for modeling power with nonnegativity cones by considering a simple example. For any function f defined on K, we have

$$\inf\{f(x) \: : \: x \text{ in } K\} = \sup\{\gamma \: : \: f(x) - \gamma \geq 0 \text{ for all } x \text{ in } K\}.$$

The problem on the left is nonconvex in general, and yet the problem on the right can be stated as a convex program with nonnegativity cones. Of course – this reformulation is only a formalism. We cannot hope to solve optimization problems involving nonnegativity cones any more than we can hope to solve general nonconvex optimization problems. The utility of this perspective is that it leads us to ask the following question.

> *What can we do by approximating a nonnegativity cone with another convex set that is amenable to off-the-shelf convex optimization solvers?*

It turns out that we can do a great deal. An especially rich literature has grown around the use of standard convex optimization frameworks to produce *certificates of nonnegativity* for various classes of functions. In the field of optimization itself, this approach has been used to devise convex relaxation frameworks for nonconvex optimization problems [1–3]. In control theory, nonnegativity certificates have been used to automate the search for Lyapunov functions [2, 4]. In statistics, these methods can be used to perform regression with shape constraints such as concavity or monotonicity [5, 6]. Nonnegativity certificates even have broader applications in analyzing and solving nonlinear partial differential equations [7–9].

The above applications were first developed with *sums of squares* or *SOS* certificates, which use the simple fact that a square of a real polynomial is evidently nonnegative. The SOS approach possess an intrinsic connection to algebra that is the source of much of its practical effectiveness. For example, powerful representation results in real algebraic geometry and functional analysis show that under appropriate regularity conditions, the existence of a suitable SOS certificate is actually necessary for a polynomial to be positive on a given domain [10, 11].

Still, SOS has its limitations. It has been shown that if $d > 2$, then as we increase the number of variables n, only a vanishing fraction of nonnegative n-variate degree-d polynomials actually admit SOS decompositions [12]. In fact, the larger the degree, the faster the gap between SOS polynomials and nonnegative polynomials grows. This is compounded by the fact that the sizes of the convex programs required by SOS methods grow exponentially with polynomial degree. In practice, we can expect the best performance out of SOS methods precisely when working with problems of modest size – either the problem should have relatively few variables or the polynomials should be of very low degree.

This thesis studies nonnegativity certificates that are well suited to problems that fall outside of SOS' favorable regime. We begin by turning our attention to *generalized polynomials* with arbitrary real exponents. These are functions of the form

$$\boldsymbol{t} \mapsto \sum_{\alpha \in \mathcal{A}} c_\alpha t_1^{\alpha_1} \cdots t_n^{\alpha_n} \quad \text{for some finite} \quad \mathcal{A} \subset \mathbb{R}^n$$

and real $(c_\alpha)_{\alpha \in \mathcal{A}}$. When considering these functions, the only reasonable measure of "complexity" is the number of variables n and the number of terms $|\mathcal{A}|$. If we were to compare this situation to polynomials, then we would say that we measure complexity by *sparsity* in the monomial basis.[1] Note however that because we are interested in nonnegativity problems, and because the exponents can be arbitrary real numbers, we must restrict the variable $\boldsymbol{t}$ positive orthant to ensure these functions are real-valued. Subject to this restriction, we can apply the substitution $t_i = \exp x_i$ to obtain

$$\boldsymbol{x} \mapsto \sum_{\alpha \in \mathcal{A}} c_\alpha \exp(\alpha_1 x_1 + \cdots + \alpha_n x_n).$$

We call such functions *signomials*.

At a technical level, signomial nonnegativity is the primary subject of this thesis. We make no assumptions whatsoever on the exponents $\mathcal{A}$, and so our methods have

---

[1]Strictly speaking this comparison doesn't hold water, since the monomial basis for these generalized polynomials cannot be graded by degree. But we have to start somewhere.

applications to "proper" signomial models with fractional and negative exponents. The more readily appreciable impact of our work, however, can be found in its consequences for polynomial nonnegativity problems upon specializing to $\mathcal{A} \subset \mathbb{N}^n$. It is easy to imagine how signomial techniques can be applied to polynomial models when working over the positive orthant. Our methods are distinguished by the fact that they apply to *global* polynomial nonnegativity problems. In particular,

> *The computational complexity of our methods for polynomials scales only with the number of terms $|\mathcal{A}|$ and is independent of both the number of variables in the polynomial and the polynomial's degree.*

This complexity scaling profile is possible because we do not take algebraic techniques as our starting point. Rather, we start with the *sums of arithmetic-geometric exponentials* or *SAGE* methodology for signomial nonnegativity, as introduced by Chandrasekaran and Shah [13]. There are two ways to certify the nonnegativity of an elementary summand in a SAGE decomposition. The namesake approach involves appeal to the arithmetic-geometric mean inequality, which has been studied as a tool for polynomial nonnegativity since 1989 [14]. A second approach – which is the ultimate source of this method's efficacy – involves an appeal to convex duality.

This thesis makes contributions on three fronts. First, we present structural properties of SAGE signomials in their original formulation and extend SAGE to the global polynomial nonnegativity problem. Second, we develop a generalization of SAGE for constrained signomial nonnegativity problems, and we undertake a multi-faceted analysis of this methodology with contributions in convex geometry, algebraic geometry, and functional analysis. Third, we provide a comprehensive implementation of our methods in an open-source software package.

The remainder of this chapter provides additional motivation for studying signomials and describes our contributions in detail.

1.1 Signomials in a broader context

The literature on signomials is quite fragmented, owing to a wide range of conventions used for this class of functions across fields and over time. In analysis of biochemical reaction networks, signomials are often called *generalized polynomials* [15, 16], or simply "polynomials over the positive orthant" [17]. In amoeba theory and tropical geometry one usually calls signomials *exponential sums* [18–20], and certain authors adjacent to these communities have used this term when writing

about SAGE [21–23]. Much of the earlier optimization literature referred to signomial programming as "generalized geometric programming," but this term now means something quite different [24].

A recurring theme across many of these works is to consider signomials in the generalized polynomial formulation. Our use of the term "signomial" to mean linear combinations of functions $x \mapsto \exp\langle\alpha, x\rangle$ is in the minority. We believe our terminology is still appropriate because numerical methods for analyzing signomial models routinely use the exponential parameterization (this is especially true in optimization). At a deeper mathematical level, we feel that the exponential formulation is somehow more "coherent" than the generalized polynomial formulation. The exponential formulation allows for affine changes to the coordinate system, it makes clear that signomials are not closed under composition, and it makes us take seriously the restriction that $\exp x = t$ is actually positive rather than merely nonnegative.

In what follows we review examples and history of signomials in mathematical modeling.

1.1.1 Dynamical systems

The nonnegativity certificates developed in this thesis have direct applications to many kinds of dynamical systems analyses. We did not have a chance to deeply explore these applications ourselves, but there are several promising opportunities moving forward.

The biggest opportunities concern systems obeying mass action dynamics [25, 26] and chemical reaction network theory [27, 28]. These systems are governed by polynomial vector fields $\dot{x}(t) = F(x(t))$ that are only defined on the positive orthant $x(t) \in \mathbb{R}^n_{++}$. Important properties of these systems can be stated in terms of the polynomial $p(z) = \det \operatorname{Jac} F(z)$ never taking the value zero on $z > 0$. Therefore one can try to certify either $p(z) > 0$ or $p(z) < 0$ over the positive orthant. These polynomials are extremely high degree and completely unapproachable by SOS methods. We expect that SAGE should perform well on these problems, partly because an earlier SAGE-like method was invented for the express purpose of analyzing chemical reaction networks and biological systems [17, 29]. An inefficient approach to SAGE known as *sums of nonnegative circuits* has also successfully been applied to biological systems analysis [30]. In Section 4.3 we show how to efficiently use SAGE for this kind of analysis.

We refer the reader to [15, 16] and [31] for other papers on these systems.

1.1.2 Signomial optimization

To the best of our limited knowledge, the study of signomials as a modeling tool actually began with optimization. The story begins with a 1961 paper by the physicist Clarence Zener, wherein he explained that minimizing very simple *posynomials* (signomials with nonnegative coefficients) reduces to solving a system of linear equations [32]. Zener suggested the tool was useful in solving certain engineering design problems and wrote a follow-up article the next year [33].

The next big leap came in 1967, when Duffin, Peterson, and Zener introduced a framework of *geometric programming* [34]. Here, one minimizes a posynomial subject to upper bound constraints on posynomials. Geometric programs are "nice" because local minima are global minima. Whether or not geometric programs are convex depends on whether you use the generalized polynomial formulation (which is not convex) or the exponential formulation (which is convex).

The term *signomial* was coined in a 1970 technical report by Duffin and Peterson, which explored the theory of geometric programming without the "posynomial" assumptions.[2] Signomial optimization is computationally intractable in the formal sense of NP-hardness. The early signomial programming literature does not emphasize this fact, probably because the concept of NP-hardness barely existed at the time (and NP-hardness of continuous optimization problems was not established until 1987 [36]).

The first polynomial-time algorithm for geometric programming came in 1994. It required geometric programs to be stated with exponential functions and used the machinery of self-concordant barrier functions [37]. The interior-point revolution that swept the optimization community brought renewed interest to geometric programming. Some influential works from the early 2000s include Chiang's monograph on applications of geometric programming in communications networks [38] and the extensive survey by Boyd, Kim, Vandenberghe, and Hassibi [24].

Nonconvex signomial optimization is a popular framework in engineering design. The historical record shows early applications in chemical engineering [39–42] followed by civil and structural engineering [43–46]. Applications in electrical engineering picked up in the mid 2000s. See, for example, [47–49]. Most recently there has been a tremendous amount of interest in signomial programming for

---

[2]The technical report was published three years later in [35].

aircraft design. For the period of 2018-2019 alone, we have [50–55]. Just this year, Virgin Hyperloop released a paper describing how solving highly-structured signomial programs in thousands of variables is a key component of their design process [56]. The conceptual source of these signomial models in engineering is the simple practice of modeling systems with non-polynomial power laws [57, 58].

1.2 This thesis

Our contributions span four papers [59–62]. We have split [60] into three chapters: one that explains a simple yet significant generalization of SAGE signomials, one that explains the analogous advance for polynomials, and one that reviews our software and describes some of our early computational experiments. The first and third of these split chapters include new content that did not appear in [60].

1.2.1 Chapter 2: Preliminaries

Here we introduce essential concepts in convex analysis, nonnegativity certificates, and notation for the rest of the thesis.

We start by introducing *convex cones* as convex sets that are closed under positive scaling. We define the arithmetic-geometric exponentials or *AGE functions* as the elementary nonnegative summands in SAGE decompositions, and we explain how the recognition problem for AGE functions can be accomplished through either of two ways: an arithmetic-geometric mean argument or a convex duality argument involving the relative entropy function.

Once we have a concrete example of nonnegativity certificates in mind, we formalize the idea of using nonnegativity cones to losslessly convexify nonconvex optimization problems. We provide a dual point of view where the convexification is accomplished with *moment cones* and we explain how global nonnegativity certificates are typically extended to constrained problems.

This chapter includes tutorial remarks on convex optimization. We comment on how the *relative entropy programs* used in SAGE computations compare to the *semidefinite programs* used in SOS computations.

1.2.2 Chapter 3: Newton polytopes and relative entropy optimization [59]

This chapter presents structural results for cones of SAGE signomials induced by a given finite set of exponents $\mathcal{A} \subset \mathbb{R}^n$. We characterize the extreme rays (conceptually, "edges") of these cones in terms of the geometry of $\mathcal{A}$ and we prove a simple

sparsity preservation property of SAGE decompositions with important theoretical and practical consequences. We leverage these structural properties to determine conditions under which signomial nonnegativity is equivalent to the existence of a SAGE decomposition. Or, what is the same, we determine conditions for when SAGE-based relative entropy programs can be used to solve a nonconvex signomial program exactly. After proving our main signomial results, we direct our machinery towards the topic of globally nonnegative polynomials. Our proposed *SAGE polynomials* provide efficient methods for certifying global polynomial nonnegativity, with complexity independent of polynomial degree.

1.2.3 Chapter 4: A new approach to constrained signomial nonnegativity [60]

Here we show that the convex duality argument behind globally nonnegative SAGE signomials can be extended to address nonnegativity over any convex set $\mathsf{X} \subset \mathbb{R}^n$. We call this broader concept *conditional SAGE* by consideration to a dual perspective and moment problems in functional analysis. The conditional SAGE approach is remarkable because the associated certificates of nonnegativity inherit tractability from the convex set X. In addition, this class of functions is completely independent from any representation of X, which is a radical departure from the situation with SOS approaches to constrained polynomial nonnegativity. We present examples of the conditional SAGE approach in signomial optimization and chemical reaction network theory. The examples illustrate how conditional SAGE can certify nonnegativity over nonconvex sets, in a process known as partial dualization.

1.2.4 Chapter 5: Sublinear circuits and signomial nonnegativity [61]

This chapter undertakes a structural analysis of conditional SAGE signomials. Towards this end, we introduce X-*circuits of a finite subset* $\mathcal{A} \subset \mathbb{R}^n$. These objects generalize the simplicial circuits of the affine-linear matroid induced by $\mathcal{A}$ to a constrained setting, by requiring a local, orthant-wise, strict sublinearity condition for the support function of $-\mathcal{A}\mathsf{X}$. These *sublinear circuits* have rich combinatorial properties when X is polyhedral, in which case X-circuits generate the one-dimensional cones of certain polyhedral fans. The X-circuit framework leads not only to generalizations of our extreme ray results for ordinary SAGE from Chapter 3, but also to substantially stronger results for ordinary SAGE itself! While working towards this chapter's main theorem we develop a duality theory for X-circuits with connections to the geometry of sets that are convex according to the geometric mean.

1.2.5 Chapter 6: An algebraic approach to signomial optimization [62]

The study of polynomial nonnegativity has benefited greatly from the tools of real algebraic geometry and functional analysis. In this chapter we show those tools can be extended to signomials in full generality.

Every finite $\mathcal{A} \subset \mathbb{R}^n$ that contains the origin gives us a real signomial ring generated by the basis functions $\{x \mapsto \exp\langle \alpha, x \rangle\}_{\alpha \in \mathcal{A}}$. Using this concept, we show if a signomial is positive on a compact set defined by convex constraints and possibly-nonconvex signomial inequalities, then there exists a conditional SAGE certificate that proves its nonnegativity. Such a result is called a *Positivstellensatz* in real algebraic geometry and it can be understood as providing *arbitrarily strong inner-approximations* of signomial nonnegativity cones. We use this Positivstellensatz to develop a practical hierarchy of relative entropy programming relaxations for approaching the value of a signomial minimization problem from below.

We also develop an elementary signomial moment theory. Our basic ingredients are a signomial Riesz-Haviland theorem and a moment-determinacy result for representing measures of signomial moment sequences. These results are combined to develop *arbitrarily strong outer-approximations* of signomial nonnegativity cones. Using these outer-approximations, we present a hierarchy of relative entropy programming relaxations for approaching the value of a signomial minimization problem from above.

1.2.6 Chapter 7: Constrained polynomial nonnegativity [60]

Chapters 4 through 6 develop the idea and theory of conditional SAGE certificates from signomial nonnegativty. Here we take that idea full circle by combining it with the "SAGE polynomials" from Chapter 3. Specifically, we show that if X is contained in $\mathbb{R}^n_+$ or invariant under reflection about the hyperplanes $\{x : x_i = 0\}$, then a representation of *conditional SAGE polynomials* that are nonnegative on X can be obtained by reduction to conditional SAGE signomials. For the cases that we consider, a given cone of conditional SAGE polynomials is tractable whenever the logarithm $\log \mathsf{X} := \{y : \exp y \in \mathsf{X}\}$ is a tractable convex set. Such sets X are "convex according to the geometric mean" in the sense of Chapter 5. Two worked examples demonstrate the efficiency of conditional SAGE polynomials relative to SOS methods for polynomial optimization.

1.2.7 Chapter 8: The `sageopt` python package [60]

This chapter describes `sageopt`: a comprehensive python package for all things SAGE. We developed this package in parallel with our mathematical work from earlier chapters.

`Sageopt` includes a subsystem for symbolic representations of signomials and polynomials, a customized convex optimization modeling system called `coniclifts`, and predefined hierarchies of convex relaxations for signomial and polynomial optimization using SAGE-based methods. `Coniclifts` is `sageopt`'s backbone. Its original purpose was to parse signomial and polynomial constraints and construct the sets "X" for use in conditional SAGE computations. It has since grown to manage all transformations between high-level SAGE constraints and low-level relative entropy programming solvers. These transformations are implemented with various presolve features based on our theoretical results. `Coniclifts` employs novel data structures so symbolic optimization modeling can be carried out using the NumPy `ndarray` datatype, which is at the core of python's scientific computing stack.

This chapter also presents some computational experiments conducted with `sageopt` from [60]. With the aim of maximizing reproducibility, we conduct these experiments using `sageopt`'s feature of independently supporting primal and dual SAGE relaxations (which is distinguished relative to available software for SOS methods). Beyond the experiments here, researchers at Aalto University, ETH Zürich, and ABB have used `sageopt` to design optimal power conversion protocols, which required solving polynomial optimization problems at scales that were impossible with SOS methods [63, 64].

Chapter 2

PRELIMINARIES

Convex optimization and certificates of function nonnegativity are the two biggest players in this thesis. We work with both of these ideas through the framework of convex cones, which might be regarded as a third player in their own right. This chapter is mostly dedicated to providing background on these three ideas along with notation for the rest of the thesis. At the end of this chapter we state and prove two technical results in convex analysis.

2.1 Vector spaces, signomials, and exponent sets

Sets generally appear as capital letters in sans-serif font. For a finite set S, we use $\mathbb{R}^{\mathsf{S}}$ to denote the set of real $|\mathsf{S}|$-tuples indexed by $s \in \mathsf{S}$. Vectors and matrices are written in boldface while their components appear in plain type. A component v_s of a vector $\boldsymbol{v} \in \mathbb{R}^{\mathsf{S}}$ can be dropped by writing $\boldsymbol{v}_{\setminus s}$. We use $\{\boldsymbol{\delta}_s\}_{s \in \mathsf{S}}$ to denote the standard basis in $\mathbb{R}^{\mathsf{S}}$ and $\Delta_{\mathsf{S}} = \{\boldsymbol{v} \in \mathbb{R}_+^{\mathsf{S}} : \langle \mathbf{1}, \boldsymbol{v} \rangle = 1\}$ to denote the probability simplex in $\mathbb{R}^{\mathsf{S}}$ (where $\mathbf{1}$ is the vector of all ones).

Let $\mathcal{A} \subset \mathbb{R}^n$ be finite. A signomial f is *supported on* $\mathcal{A}$ if it can be written as

$$f(\boldsymbol{x}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \exp\langle \alpha, \boldsymbol{x} \rangle$$

for a vector $\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}}$. To keep notation compact, we sometimes write $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ in reference to the signomial above. We also use $\mathrm{e}^\alpha : \mathbb{R}^n \to \mathbb{R}$ for basis functions $\mathrm{e}^\alpha(\boldsymbol{x}) = \exp\langle \alpha, \boldsymbol{x} \rangle$, so that signomials can be specified by $f = \sum_{\alpha \in \mathcal{A}} c_\alpha \mathrm{e}^\alpha$. This latter notation for strikes a nice balance between allowing us to examine individual terms "$c_\alpha \mathrm{e}^\alpha$" and writing signomials out in full.

Signomial (and polynomial) exponent vectors $\mathcal{A}$ are regarded *both as sets and as linear operators*. The forward operator is

$$\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^{\mathcal{A}} \qquad \boldsymbol{x} \mapsto (\langle \alpha, \boldsymbol{x} \rangle)_{\alpha \in \mathcal{A}}$$

and the adjoint is

$$\mathcal{A}^\dagger : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^n \qquad \boldsymbol{y} \mapsto \sum_{\alpha \in \mathcal{A}} \alpha y_\alpha.$$

If this abstraction creates any confusion, then one should simply think of $\mathcal{A}$ as a matrix with rows $\alpha \in \mathbb{R}^n$.

2.2 Convex cones

We call a set C *convex* if it contains all of its line segments. That is, C is convex if for every $x, y \in C$ and $\lambda \in [0, 1]$, we have $\lambda x + (1 - \lambda)y \in C$. The convex hull of a set K is the smallest convex set that contains K; this set is denoted by "conv K."

Convex cones are among the nicest of convex sets. They are the convex sets C that additionally satisfy

$$x \in C \quad \Rightarrow \quad \lambda x \in C \quad \text{for all } \lambda > 0.$$

Familiar convex cones include the nonnegative orthant and the set of symmetric positive semidefinite matrices of a given order. Another familiar (albeit generate) class of convex cones is given by linear spaces. With the aim of excluding such degenerate cases, we shall call a cone *pointed* if it contains no lines; the nonnegative orthant is the quintessential example of a pointed convex cone.

2.2.1 Conic hulls and extreme rays

The conic hull of $K \subset \mathbb{R}^S$ is the set formed by adjoining the origin to the smallest convex cone containing K. We denote this by co K and observe that it can be expressed as co $K = (\cup_{\lambda > 0} \lambda K) \cup \{0\}$. Next, we define an *extreme ray* of a convex cone C as a ray R that is contained in C and that satisfies the following condition: if an open line segment in C intersects R, then the line segment is contained entirely within R.

Extreme rays are important because if C is a closed and pointed convex cone, then it is the conic hull of any point set that generates all of its extreme rays. Moreover, the union of all extreme rays of C is the smallest set K for which $C = \text{conv } K$. As one example of this phenomenon, the extreme rays of nonnegative orthant $\mathbb{R}^S_+$ are the rays induced by standard basis vectors $\{\delta_s\}_{s \in S}$. As another example, the extreme rays of the cone $\{(x, t) : \|x\|_2 \leq t\}$ are the rays generated by vectors $(x, \|x\|_2)$ for nonzero x. For technical discussion we refer the reader to [65, Chapter 18].

2.2.2 Dual cones

The main advantage in thinking in terms of convex cones is the idea of conic duality. Formally, to any convex cone $C \subset \mathbb{R}^S$, we associate the *dual cone*

$$C^\dagger := \{y \in \mathbb{R}^S : \langle x, y \rangle \geq 0 \text{ for all } x \text{ in } C\}.$$

The original set C is analogously called the *primal cone*. There are some situations where the dual cone is obvious from the primal (for example, when $C = \mathbb{R}^S_+$), but

usually some calculations are involved in determining the dual cone.

There are several facts one can use to reason about a dual cone without going through the full conic duality calculations. One such fact is that if $C \subset C'$, then $(C')^\dagger \subset C^\dagger$; this is known as *order-reversal*. The most important fact is that taking dual cones is an involution, up to closure: $(C^\dagger)^\dagger = \text{cl } C$. In this thesis, we often cite the fact that if $C = \sum_{\ell \in [m]} C_\ell$ is a Minkowski sum[1] of convex cones C_ℓ, then $C^\dagger$ is the intersection of the associated dual cones: $C^\dagger = \cap_{\ell \in [m]} C_\ell^\dagger$.

2.2.3 Nonnegativity cones and moment cones

Suppose we have a family of real-valued basis functions $\phi = (\phi_i)_{i \in [m]}$ defined on an abstract set S. These data induce a finite-dimensional closed convex cone

$$C = \left\{ c \in \mathbb{R}^m \ : \ \sum_{i \in [m]} c_i \phi_i(x) \geq 0 \text{ for all } x \text{ in } S \right\}.$$

Such sets C are usually intractable. However, they are a powerful formalism that show how many problems in applied mathematics can nominally be approached through convex optimization. We can unlock methods to solve such problems approximately by developing tractable approximations C' that stand in for C. This thesis is particularly concerned with *inner approximations* $C' \subset C$.

The nonnegativity cone formalism is amenable to the machinery of conic duality. Recalling that $(C^\dagger)^\dagger = C$, one can deduce

$$C^\dagger = \text{cl co}\{ (\phi_1(x), \ldots, \phi_m(x)) \ : \ x \in S \},$$

where "co" is the conic hull operator that we defined in Subsection 2.2.1. One can gain intuition for $C^\dagger$ by considering the *moment body*

$$M = \text{conv}\{ (\phi_1(x), \ldots, \phi_m(x)) \ : \ x \in S \}.$$

The term "moment" here references moments in probability theory, since the operation of taking a convex hull can be framed as taking all vector-valued expectations $\mathbb{E}_{x \sim F} \phi(x)$ where F is a probability measure supported on S. The particular connection between the moment body and $C^\dagger$ is that $C^\dagger = \text{cl} (\cup_{\lambda > 0} \lambda M)$. It is therefore standard to call $C^\dagger$ a *moment cone*.

---

[1]The Minkowski sum of sets A, B in a shared vector space is $A + B = \{a + b \ : \ a \in A, b \in B\}$.

2.3 Sums of arithmetic-geometric exponentials (SAGE)

This thesis studies nonnegative signomials, nonnegative polynomials, and associated nonconvex optimization problems. However, our contributions in the polynomial realm are fairly direct consequences of our contributions in the signomial realm. Since *sums of arithmetic-geometric exponentials* [13] are our starting point for understanding nonnegative signomials, we describe that approach here.

The simplest class of nonnegative signomials are the *posynomials*: the signomials with all nonnegative coefficients. An *arithmetic-geometric exponential* (or *AGE function*) is almost a posynomial. These functions are, by definition, the globally nonnegative signomials $f(x) = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha(x)$ with at most one negative term. A signomial is *SAGE* if it is a sum of AGE functions.

There are two ways to prove nonnegativity of an AGE function. The namesake approach involves an appeal to the arithmetic-geometric mean inequality (or AM/GM inequality). A second approach, which is the ultimate source of this method's efficacy and deeper theory, involves relative entropy certificates obtained by convex duality.

2.3.1 A sketch of the arithmetic-geometric approach

In its usual form, the AM/GM inequality says that for every $t \in \mathbb{R}_{++}^m$ and λ in the interior of the probability simplex Δ_m, we have $t^\lambda \leq \langle \lambda, t \rangle$. To see how this relates to signomials, we can consider how the AM/GM inequality is equivalent to convexity of the exponential function. That is, $\exp\langle \lambda, y \rangle \leq \langle \lambda, \exp y \rangle$ for every $y \in \mathbb{R}^m$ and $\lambda \in \Delta_m$. With this convexity observation in the back of our mind, we fix a set of exponent vectors $\mathcal{A} \subset \mathbb{R}^n$, pick a weighting vector $\lambda \in \operatorname{int} \Delta_\mathcal{A}$, and set $\beta = \mathcal{A}^\dagger \lambda$. Choosing β in this way simply means that β is in the convex hull of $\mathcal{A}$ and that the claim of membership "$\beta \in \operatorname{conv} \mathcal{A}$" is certified by λ. From here we can deduce a chain of identities

$$\mathrm{e}^\beta(x) = \exp\langle \mathcal{A}^\dagger \lambda, x \rangle = \exp\langle \lambda, \mathcal{A}x \rangle \leq \langle \lambda, \exp \mathcal{A}x \rangle = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha \mathrm{e}^\alpha(x)$$

which tell us that $\sum_{\alpha \in \mathcal{A}} \lambda_\alpha \mathrm{e}^\alpha + c_\beta \mathrm{e}^\beta$ is nonnegative on $\mathbb{R}^n$ whenever $c_\beta \geq -1$. This particular construction is specialized, but it provides most of the idea for using the AM/GM inequality as a way to recognize AGE functions.

2.3.2 Relative entropy certificates from convex duality

The *relative entropy function* is the continuous extension of

$$D(\boldsymbol{u}, \boldsymbol{v}) = \sum_{s \in S} u_s \log\left(\frac{u_s}{v_s}\right)$$

to the product of nonnegative orthants $(\boldsymbol{u}, \boldsymbol{v}) \in \mathbb{R}_+^S \times \mathbb{R}_+^S$. Relative entropy is convex when viewed as a function of a single concatenated argument $\boldsymbol{w} = (\boldsymbol{u}, \boldsymbol{v})$; this property is called *joint convexity* and is stronger than being convex in one argument while the other is fixed.

Let $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ be a signomial with $c_{\setminus\beta} \geq \boldsymbol{0}$. The problem of deciding nonnegativity of f takes two steps. First, since $e^\beta(\boldsymbol{x}) > 0$ for all $\boldsymbol{x}$, we can divide out this basis function to obtain $g = c_\beta + \sum_{\alpha \in \mathcal{A} \setminus \beta} c_\alpha e^{[\alpha - \beta]}$ without affecting nonnegativity. Second, because g is a posynomial plus a constant, bounds on its minimum can be certified through the principle of strong duality in convex optimization. The outcome of this duality argument is that f is nonnegative if and only if there exists a $\boldsymbol{v} \in \mathbb{R}^{\mathcal{A}}$ where (i) $\langle \boldsymbol{1}, \boldsymbol{v} \rangle = 0$, (ii) $\mathcal{A}^\dagger \boldsymbol{v} = \boldsymbol{0}$, and (iii) $D(\boldsymbol{v}_{\setminus\beta}, e c_{\setminus\beta}) \leq c_\beta$. The joint convexity of relative entropy means it is possible to efficiently optimize over the cone of AGE functions for any given $\boldsymbol{\beta} \in \mathcal{A}$.

2.4 Optimization with nonnegativity and moment cones

Here we show how any nonconvex minimization problems can be reformulated exactly into a convex optimization problem involving a nonnegativity cone. We also construct an equivalent convex problem stated in terms of a moment cone. We discuss how bounds on these problems are obtained by working with inner-approximations of nonnegativity cones and outer-approximations of moment cones.

Throughout, we consider the space of functions spanned by linear combinations of components in a vector-valued map $\phi = (\phi_1, \ldots, \phi_m)$. We assume that $\phi_1(\boldsymbol{x}) = 1$ for all $\boldsymbol{x}$ in $\mathbb{R}^n$. We use C to denote the cone of vectors $\tilde{\boldsymbol{c}}$ where $\boldsymbol{x} \mapsto \langle \tilde{\boldsymbol{c}}, \phi(\boldsymbol{x}) \rangle$ is globally nonnegative, and we would like to compute the infimum of $f(\boldsymbol{x}) = \langle \boldsymbol{c}, \phi(\boldsymbol{x}) \rangle$.

2.4.1 Using nonnegativity cones and inner-approximations

The idea of using nonnegativity certificates begins with the simple observation that

$$f_{\mathbb{R}^n}^\star := \inf_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) = \sup\{\gamma \ : \ f - \gamma \text{ is nonnegative on } \mathbb{R}^n\}.$$

Stating this with nonnegativity cones, we have

$$f_{\mathbb{R}^n}^\star = \sup\{\gamma \ : \ \boldsymbol{c} - \gamma \delta_1 \text{ in } \mathsf{C}\}. \tag{2.1}$$

The constraint in Problem (2.1) simply says that the coefficient vector of $f - \gamma$ with respect to ϕ belongs to the nonnegativity cone induced by ϕ.

The problem (2.1) is convex, but since it encodes an essentially arbitrary optimization problem, it stands to reason that it is intractable. We really introduced this problem so we can consider what happens when we replace C by a smaller cone K. Regardless of the precise relationship between K and C, it is clear that $\mathsf{K} \subset \mathsf{C}$ implies

$$f_{\mathbb{R}^n}^{(\mathsf{K})} = \sup\{\gamma \ : \ \boldsymbol{c} - \gamma\boldsymbol{\delta}_1 \text{ in } \mathsf{K}\} \leq f_{\mathbb{R}^n}^\star. \tag{2.2}$$

For example, C might represent the cone of globally nonnegative signomials supported on exponents $\mathcal{A} \subset \mathbb{R}^n$, and K might denote the cone of SAGE signomials supported on $\mathcal{A}$.

2.4.2 Using moment cones and outer-approximations

There is another convex cone program that can be used to represent $f_{\mathbb{R}^n}^\star$ – *the moment relaxation*. To obtain this problem we begin by convincing ourselves that $f_{\mathbb{R}^n}^\star = \inf\{\langle \boldsymbol{c}, \boldsymbol{v} \rangle \ : \ \boldsymbol{v} \in \mathsf{M}\}$, where $\mathsf{M} = \operatorname{conv} \phi(\mathbb{R}^n)$ is the moment body associated with the basis functions $(\phi_1, \ldots, \phi_m)$. To reformulate this using the moment cone rather than the moment body, recall the relationship $\mathsf{C}^\dagger = \operatorname{cl}(\cup_{\lambda>0}\lambda\mathsf{M})$. From this it is clear that $\mathsf{M}' = \{\boldsymbol{v} \ : \ \boldsymbol{v} \in \mathsf{C}^\dagger, \ \langle \boldsymbol{\delta}_1, \boldsymbol{v} \rangle = 1\}$ is at least as large as M. In fact, the set M' can only differ from M up to closure, and we have

$$f_{\mathbb{R}^n}^\star = \inf\{\langle \boldsymbol{c}, \boldsymbol{v} \rangle \ : \ \langle \boldsymbol{\delta}_1, \boldsymbol{v} \rangle = 1, \ \boldsymbol{v} \text{ in } \mathsf{C}^\dagger\}. \tag{2.3}$$

This process of arriving at the moment relaxation is elementary, but somewhat cumbersome. Later in this chapter we explain the framework of duality for convex cone programs that makes the transition from (2.1) to (2.3) effortless.

If we suppose once more that we we replace C by a smaller cone K, then the order-reversal property of dual cones tells us that $\mathsf{C}^\dagger \subset \mathsf{K}^\dagger$. Thus following identity holds under generic conditions

$$f_{\mathbb{R}^n}^{(\mathsf{K})} = \inf\{\langle \boldsymbol{c}, \boldsymbol{v} \rangle \ : \ \langle \boldsymbol{\delta}_1, \boldsymbol{v} \rangle = 1, \ \boldsymbol{v} \text{ in } \mathsf{K}^\dagger\}. \tag{2.4}$$

This is very useful, since optimal solutions problems to like (2.4) often contain information about locations of f's minimizers.

2.4.3 Constrained nonnegativity problems

So far we have had C be a cone of functions nonnegative on $\mathbb{R}^n$. We made this restriction to provide a clear connection to the SAGE certificates of global signomial

nonnegativity from Section 2.3. But as Subsection 2.2.3 meant to suggest, it is fine to have C be a cone of functions that are nonnegative on *any abstract set*.

The bigger picture here is that any means of certifying nonnegativity over a set S gives us two convex problems – one with a nonnegativity cone and one with a moment cone – that we can use to bound or even solve nonconvex optimization problems. A key question then becomes how to extend certificates defined for one set S (particularly S = $\mathbb{R}^n$) to another smaller set S' ⊂ S. We touch upon this question twice in Chapter 3, and then we investigate it seriously in Chapter 4.

2.5   Background on cone programming

Convex optimization is the minimization of a convex function f over a convex set X. In this thesis, we usually assume f is linear and X is described as the intersection of a convex cone with an affine subspace; this class of problems is known as linear cone programs (or simply cone programs). Any convex program can be written as a cone program a modest increase in dimension. In fact, the express purpose of software such as CVXPY and Yalmip is simply to manage transformations between user-specified models and solver-mandated standard forms.

2.5.1   Semidefinite and relative entropy programming

Suppose our feasible set is X = $\{\boldsymbol{x} \ : \ \boldsymbol{Ax} + \boldsymbol{b} \in \mathsf{K}\}$ for a convex cone K and a linear operator $\boldsymbol{A}$. If K is the cone of positive semidefinite matrices, then we could call the cone program a *semidefinite program* or an *SDP*. This class of problems has proven extremely powerful from a modeling perspective and (per a result by Shor [1]) is the computational basis for working with sums of squares certificates of polynomial nonnegativity. SDPs have been studied intensely since the 1990s, but remain difficult to solve at large scales. The underlying cause for the difficulty of SDP is simply a matter of linear algebra: unless your problem has special structure, the search directions computed by optimization algorithms require solving large-scale dense systems of linear equations [66].

This thesis focuses on *relative entropy programming* or *REP*, because of its importance for SAGE nonnegativity certificates. The epigraph of relative entropy

$$\mathsf{C_S} = \{(\boldsymbol{u}, \boldsymbol{v}, w) \in \mathbb{R}_+^{\mathsf{S}} \times \mathbb{R}_+^{\mathsf{S}} \times \mathbb{R} \ : \ D(\boldsymbol{u}, \boldsymbol{v}) \leq w\}$$

is a cone, because $(\boldsymbol{u}, \boldsymbol{v}) \mapsto \sum_{s \in \mathsf{S}} u_s \log(u_s/v_s)$ is convex and homogeneous of degree one. REPs have theoretically been tractable since the introduction of self-concordant barrier functions in 1994 [37], however methods for large scale REP

have only recently been developed. The search directions computed by optimization algorithms involve large scale *sparse* systems of linear equations. Provided one uses a reliable solver (e.g., MOSEK 9), it is entirely possible to solve REPs involving C_S with $|S| \approx 10^6$, even on a personal laptop.

2.5.2 Conic duality

Suppose that we want to minimize a linear function $x \mapsto \langle c, x \rangle$ over the set $\{x : Ax = b, x \in K\}$, where K is a convex cone, A is a linear operator from $\mathbb{R}^R$ to $\mathbb{R}^S$, and b is a vector in $\mathbb{R}^S$. We take this data to define a *primal problem*

$$\mathrm{Val}_{\inf} = \inf\{\, \langle c, x \rangle \,:\, Ax = b,\, x \in K \}. \tag{2.5}$$

From the primal, one may derive a *dual problem*

$$\mathrm{Val}_{\sup} = \sup\{\, \langle b, y \rangle \,:\, c - A^\dagger y \in K^\dagger \}. \tag{2.6}$$

By analogy with linear programming, if (i) x is primal-feasible, (ii) y is dual-feasible, and (iii) $\langle x, c - A^\dagger y \rangle = 0$, then (x, y) are primal-dual optimal.

The usual process of deriving the dual via a Lagrangian argument shows that $\mathrm{Val}_{\sup} \leq \mathrm{Val}_{\inf}$ always holds; this is the phenomenon of *weak duality*. The situation where $\mathrm{Val}_{\inf} = \mathrm{Val}_{\sup}$ is known as *strong duality*. It is easy to show that the aforementioned optimality conditions (i) – (iii) imply strong duality. Strong duality can also hold when one of these two problems is infeasible. For example, if the primal is infeasible and the dual is unbounded, then we still have $\mathrm{Val}_{\inf} = \mathrm{Val}_{\sup} = +\infty$.

Strong duality holds generically for cone programs, however, it can fail. If we are working with well-behaved cones like $K = \mathbb{R}^n_+$, then the only failure case is when both the primal and dual problems are infeasible ($\mathrm{Val}_{\inf} = +\infty$ and $\mathrm{Val}_{\sup} = -\infty$). For more general cones like those involving positive semidefinite matrices, it is possible that $\mathrm{Val}_{\sup} < \mathrm{Val}_{\inf}$ even when both values are finite.

2.5.3 Algorithms for cone programming

Our best algorithms for cone programming are primal-dual algorithms. These are algorithms that, given a primal-dual pair (2.5)-(2.6), generate a sequence of points (x_t, y_t) that converge to a limit satisfying optimality conditions similar to (i)–(iii). The precise optimality conditions are only slightly modified so that certificates of primal infeasibility or dual infeasibility can also be recovered.

Within primal-dual methods, there are prominent categories of *first-order algorithms* and *second-order algorithms*. Which type of algorithm is preferable depends on

one's situation. First-order algorithms have become very popular in recent years, due
to claims of superior scalability for problems arising in data science. Our experience
is that these claims do not hold up if a user needs high-quality solutions and reliable
algorithm behavior. Rather, our experience (in this thesis and elsewhere) is that one
should use a second-order solver whenever possible.

The optimization community has largely settled on a particular type of second-order
algorithms for cone programming. These algorithms are *path-following interior
point methods* featuring some kind of *Mehrotra corrector*. The Mehrotra corrector
was first famously introduced for linear programming by Sanjay Mehrotra back in
1992 [67], and was generalized to semidefinite programming through the framework
of Euclidean Jordan algebras (see [68]) in the late 1990s. Various attempts were
made to generalize the Mehrotra corrector for problems beyond semidefinite pro-
gramming, but was only really accomplished in 2019 [69, 70]. At present, MOSEK
is the only easily available solver with a Mehrotra corrector for relative entropy
programming.

2.6 Two technical notes in convex analysis

Through Section 2.5 it has become clear that for a given nonconvex optimization
problem, the nonnegativity-cone reformulation and the moment-cone reformulation
are dual to one another. This leads to a very important question: when can we
be certain that strong duality holds when we pass to our inner-approximation of a
nonnegativity cone? Here we present and prove an abstract result that ensures strong
duality holds in all reasonable settings.

The reader may skip this section if pressed for time.

Theorem 2.6.1. *Let* C *be a closed and pointed convex cone in* $\mathbb{R}^{\mathsf{S}}$, *and fix* $\boldsymbol{a} \in \mathsf{C} \setminus \{\boldsymbol{0}\}$.
For every $\boldsymbol{c} \in \mathbb{R}^{\mathsf{S}}$, *the primal-dual pair*

$$p^{\star} = \sup\{\gamma : \boldsymbol{c} - \gamma\boldsymbol{a} \ \text{in} \ \mathsf{C}\} \quad \text{and} \quad d^{\star} = \inf\{\langle \boldsymbol{c}, \boldsymbol{v} \rangle : \langle \boldsymbol{a}, \boldsymbol{v} \rangle = 1, \ \boldsymbol{v} \ \text{in} \ \mathsf{C}^{\dagger}\}$$

exhibits strong duality. I.e., $p^{\star} = d^{\star}$.

We have stated Theorem 2.6.1 in a general form because it offers a pleasant connec-
tion to eigenvalue problems.

Example 2.6.2. Let $\mathsf{S} = \{(i, j) : 1 \leq i \leq j \leq n\}$, identify $\mathbb{R}^{\mathsf{S}}$ with the space of real
symmetric matrices of order n, and choose C as the cone of positive semidefinite
matrices of order n. Given a matrix $\boldsymbol{X}$, we can represent the minimum eigenvalue

$\lambda_{\min}(X) = p^{\star}$ by setting $a := (\delta_{ij} \ : \ (i, j) \in \mathsf{S})$ (representing the identity matrix) and $c = (X_{ij} \ : \ (i, j) \in \mathsf{S})$ (representing X).

Because C is pointed and the identity matrix is positive definite, we also have $\lambda_{\min}(X) = d^{\star}$. This is really a disguised version of the variational characterization of the smallest eigenvalue! To see why, start by noting that $\mathsf{C}^{\dagger} = \mathsf{C}$ (i.e., the cone of positive semidefinite matrices is self-dual) so the dual-feasible set $\{v \in \mathsf{C}^{\dagger} \ : \ \langle a, v \rangle = 1\}$ is compact. Then we convince ourselves that the optimal solution $v^{\star}$ represents a rank-1 matrix, which we can write with an outer-product $v^{\star} \equiv u \otimes u$ for a unit vector u. This unit vector is none other than the bottom eigenvector of X.

Of course, Theorem 2.6.1 is also applicable to our convexifications of nonconvex problems.

Example 2.6.3. Let $\mathsf{S} = [m]$ and consider a vector-valued mapping $\phi : \mathsf{X} \to \mathbb{R}^m$ where $\phi_1(x) = 1$ for all x in X. From ϕ and X we construct the nonnegativity cone $\mathsf{C} = \{\tilde{c} \in \mathbb{R}^m \ : \ \langle \tilde{c}, \phi(x) \rangle \geq 0 \text{ for all } x \text{ in } \mathsf{X}\}$. Given a function $f(x) = \langle c, \phi(x) \rangle$, we can represent $p^{\star} = \inf\{f(x) \ : \ x \in \mathsf{X}\}$ by taking $a = \delta_1$.

If the coordinate functions of ϕ are linearly independent on X, then C is pointed and we have $p^{\star} = d^{\star}$. This says we can completely characterize X-nonnegative functions in the span of $\{\phi_i\}_{i \in [m]}$ by requiring that $d^{\star} \geq 0$. Moreover: strong duality will hold for any primal-dual pair where C is replaced by some $\mathsf{C}' \subset \mathsf{C}$, so long as C' contains the first standard basis vector.

We also provide a far more elementary proposition. This will be useful to us in crafting fully rigorous arguments around dual SAGE cones.

Proposition 2.6.4. *If C is a closed and pointed convex cone that contains the nonnegative orthant $\mathbb{R}^{\mathsf{S}}_+$, then $\mathsf{C}^{\dagger}$ can be expressed as $\mathsf{C}^{\dagger} = \mathrm{cl}(\mathsf{C}^{\dagger} \cap \mathbb{R}^{\mathsf{S}}_{++})$.*

2.6.1 Proof of Theorem 2.6.1

We must establish a lemma before proving the main theorem.

Lemma 2.6.5. *Fix a closed convex cone K in $\mathbb{R}^{\mathsf{S}}$. If a in $\mathsf{K}^{\dagger}$ is such that*

$$\mathsf{X} := \{x \ : \ \langle a, x \rangle = 1, \ x \text{ in } \mathsf{K}\}$$

is nonempty, then $\mathrm{cl} \, \mathrm{co} \, \mathsf{X} = \mathsf{K}$.

Proof. Certainly the conic hull of X is contained within K, and the same is true of its closure. The task is to show that every x in K also belongs to cl co X; we do this by case analysis on $b := \langle a, x \rangle$.

By the assumptions $a \in K^{\dagger}$ and $x \in K$, we must have $b \geq 0$. If b is positive then the scaling $\tilde{x} := x/b$ belongs to K and satisfies $\langle a, \tilde{x} \rangle = 1$. That is, $b > 0$ gives us $\tilde{x}$ in X. Simply undo this scaling to recover x and conclude $x \in$ co X. Now suppose $b = 0$. Here we consider the sequence of points $y_n := x_o + nx$, where x_o is a fixed but otherwise arbitrary element of X. Each point y_n belongs to K, and has $\langle a, y_n \rangle = 1$, hence the y_n are contained in X. It follows that the *scaled points y_n/n* are contained in cl co X, and the same must be true of their limit $\lim_{n \to \infty} y_n/n = x$.

Since x in K was arbitrary, we have cl co X = K. $\qquad\square$

We are now ready to prove Theorem 2.6.1.

Because C is pointed, $C^{\dagger}$ is full-dimensional, so there exists no nonzero vector $\tilde{a} \in \mathbb{R}^S$ where $\langle \tilde{a}, v \rangle = 0$ for all $v \in C^{\dagger}$. Consider this fact with $a \in C \setminus \{0\}$ to see that there exists a $v \in C^{\dagger}$ with $\langle a, v \rangle = 1$. This tells us that the dual feasible set $\{v : \langle a, v \rangle = 1, v$ in $C^{\dagger}\}$ is nonempty.

Since the dual problem is feasible, a proof that $d^\star = p^\star$ can be divided into the cases $d^\star = -\infty$, and $d^\star$ in $\mathbb{R}$. The proof in the former case is trivial; weak duality combined with $p^\star \geq -\infty$ gives $d^\star = p^\star$. In the latter case we prove $p^\star \geq d^\star$ by showing that $c^\star := c - d^\star a$ belongs to C.

To prove $c^\star \in$ C we will appeal to Lemma 2.6.5 with $K := C^{\dagger}$. Clearly the set $X = \{v : \langle a, v \rangle = 1, v$ in $K\}$ is precisely the [nonempty] feasible set for computing $d^\star$, and so from the definition of $d^\star$ we have $\langle c^\star, v \rangle \geq 0$ for all v in X. The inequality also applies to any v in cl co X, which by Lemma 2.6.5 is equal to $K^{\dagger}$. Therefore the definition of $d^\star$ ensures $c^\star$ is in $K^{\dagger}$. Using $K^{\dagger} \equiv$ C, we have the desired result.

2.6.2 Proof of Proposition 2.6.4

We begin with some definitions from Rockafellar's *Convex Analysis* [65]. The *relative interior* of a convex set C (denoted ri C) is the interior of C under the topology induced by its *affine hull*

$$\text{aff } C := \{x + t(y - x) : x, y \in C, \ t \in \mathbb{R}\}.$$

A face of C is any closed convex $F \subset C$ with the following property: if the line segment $L := \{\lambda s_1 + (1 - \lambda)s_2 : 0 \le \lambda \le 1\}$ is contained in C and the relative interior of L hits F, then the entirety of L is contained in F.

And now we prove the proposition. Rockafellar's [65, Theorem 18.2] says for any convex set T, every relatively open set contained in T is contained in the relative interior of some *face* of T. We consider that statement with $T = \mathbb{R}^S_+$ (so $\operatorname{ri} T = \operatorname{int} T = \mathbb{R}^S_{++}$). By our assumption $C^\dagger \cap \operatorname{ri} T \ne \emptyset$, the only face of T which contains $C^\dagger$ is T itself. Certainly $\operatorname{ri} C^\dagger = \operatorname{int} C^\dagger$ is relatively open, we have $\operatorname{ri} C^\dagger \subset \operatorname{ri} T$, so the claim follows by the identity $C^\dagger = \operatorname{cl} \operatorname{ri} C^\dagger$.

C h a p t e r   3

NEWTON POLYTOPES AND RELATIVE ENTROPY OPTIMIZATION

3.1   Introduction

This chapter presents the first study of SAGE certificates for signomial nonnegativity following their introduction by Chandrasekaran and Shah. We present structural results for these certificates such as a characterization of the extreme rays of SAGE cones and an appealing form of sparsity preservation. These lead to a number of important consequences such as conditions under which signomial nonnegativity is equivalent to the existence of a SAGE decomposition; our results represent the broadest-known class of nonconvex signomial optimization problems that can be solved efficiently via convex relaxation. Much of our analysis concerns a signomial's *Newton polytope*: the convex hull of its exponent vectors. We find particularly rich interactions between the convex duality underlying SAGE certificates and the face structure of Newton polytopes.

After proving our main signomial results, we direct our machinery towards the topic of globally nonnegative polynomials. This begins by making a small modification to SAGE that provides a new notion of globally nonnegative "SAGE polynomials." The complexity of working with these SAGE polynomials is, remarkably, independent of the number of variables in the polynomial or the polynomial's degree. We obtain several results on these polynomials as corollaries from our signomial results.

Analysis by Newton polytopes has a long history in the study of sparse polynomials. Prominent examples in this area include Khovanskii's fewnomials [71, 72], Reznick's agiforms [14, 73], and Bajbar and Stein's work on polynomial coercivity [74]. Many such works "signomialize" polynomials via a substitution $t_j \leftarrow \exp x_j$ in certain intermediate proofs. We adopt a different perspective, where signomials are the first-class object.

3.1.1   Chapter outline and summary of results

In Section 3.3 we prove a number of new structural properties of SAGE certificates. Theorem 3.3.1 is an important sparsity-preserving property: if a signomial f is SAGE, then there exists a decomposition $f = \sum_k f_k$ where each f_k is an AGE

function that consists only of those terms that appear in f. Furthermore, the process of summing the f_k to obtain f results in no cancellation of coefficients on basis functions $x \mapsto \exp\langle\alpha, x\rangle$. Theorem 3.3.3 goes on to provide a characterization of the extreme rays of the cone of SAGE functions; in particular, all nonatomic extreme rays are given by AGE functions that are supported on simplicial Newton polytopes.

Section 3.4 leverages the understanding from Section 3.3 to derive a collection of structural results which describe when nonnegative signomials are SAGE, with the Newton polytope being the primary subject of these theorems' hypotheses. Theorem 3.4.1 is concerned with cases where the Newton polytope is simplicial, while Theorems 3.4.2 and 3.4.3 concern when it "decomposes" in an appropriate sense. Each of these theorems exhibits invariance under nonsingular affine transformations of the exponent vectors. Corollaries 3.4.5 and 3.4.6 show how Theorem 3.4.1 applies to signomial optimization problems. We conclude the section with a result on conditions under which SAGE can recognize signomials that are bounded below (Theorem 3.4.7).

In Section 3.5 we specialize our results on signomials to polynomials, by defining a suitable "signomial representative" of a polynomial, and requiring that the signomial admit a SAGE decomposition. The resulting class of *SAGE polynomials* inherits a tractable representation from the cone of SAGE signomials (Theorem 3.5.1) as well as structural properties on sparsity preservation and extreme rays (Corollaries 3.5.5 and 3.5.6). Moving from a polynomial to a signomial representative is simple but somewhat delicate, yielding both stronger results (Corollary 3.5.2) and weaker results (Corollary 3.5.3) than in the signomial case. We then situate our results on SAGE polynomials in the broader literature, with specific emphasis on *sums of nonnegative circuits* (SONC) and *sums of squares* (SOS). The section is concluded with a discussion on how our results provide the basis for a sparsity-preserving hierarchy of convex relaxations for polynomial optimization problems.

Section 3.6 demonstrates that there are meaningful senses in which our results from Section 3.4 cannot be improved upon. Through Theorem 3.6.9, we provide a novel dual characterization of conditions under which the SAGE cone and the cone of nonnegative signomials coincide.

3.1.2   Related work: algorithms for signomial programming

In the taxonomy of optimization problems, geometric programming is to signomial optimization what convex quadratic programming is to polynomial optimization.

Current approaches to global signomial optimization use successive linear or geometric programming approximations together with branch-and-bound [75–84]. Equality constrained signomial programs are often treated by penalty or augmented Lagrangian methods [85] and are notoriously difficult to solve [86, 87]. The SAGE approach to signomial optimization does not involve branch-and-bound, and does not entail added complexity when considering signomial equations instead of inequalities.

3.1.3 Related work: sums of squares and polynomial optimization

There is a large body of work on SOS certificates for polynomial nonnegativity, and the resulting convex relaxations for polynomial optimization problems [1–3]. Over the course of this chapter we make two contributions which have direct parallels in the SOS literature.

Our results in Section 3.3 are along the lines of David Hilbert's 1888 classification of the number of variables "n" and the degrees "$2d$" for which SOS-representability coincides with polynomial nonnegativity [88]. The granularity with which we seek such a classification is distinct from that in the SOS literature, as there is no canonical method to take finite-dimensional subspaces of the infinite-dimensional space of signomials.

A principal drawback of the SOS method is that its canonical formulation requires a semidefinite matrix variable of order $\binom{n+d}{d}$ – and the size of this matrix is exponential in the degree d. In Section 3.5 we use SAGE signomials to certify polynomial nonnegativity in a way which is unaffected by the polynomial's degree. Subsection 3.5.5 compares our proposed method to SOS, as well as refinements and variations of SOS which have appeared in the literature: [89–92].

3.1.4 Related work: certifying nonnegativity via the am/gm inequality

As we explained in Section 2.3, the "AGE functions" in a SAGE decomposition may be proven nonnegative in either of two ways. The first approach is to certify a particular relative entropy inequality over a signomial's coefficients. This approach (which we describe again in Section 3.2) is known to be computationally tractable, and it provides a convenient tool for proving structural results for the set of SAGE certificates. The second method is to find weights for an appropriate am/gm inequality over a signomial's coefficients; this latter method directly connects SAGE to a larger literature on certifying function nonnegativity via the am/gm inequality, which we summarize next.

The earliest systematic theoretical studies in this area were undertaken by Reznick [14, 73] in the late 1970s and 1980s. The first developments of any computational flavor came from Pébay, Rojas, and Thompson in 2009 [93], via their study of polynomial maximization. Pébay et al. used tropical geometry and $\mathcal{A}$-discriminants to determine a complete and explicit characterization for the supremum of a polynomial over $\mathbb{R}^n_+$ (or signomial over $\mathbb{R}^n$) supported on a circuit ([93, Theorem 2.10]). In this context, a function is supported on a circuit if the monomial exponents $\{\boldsymbol{\alpha} \; : \; c_{\boldsymbol{\alpha}} \neq 0\}$ are a minimal affinely-dependent set. In 2011, Ghasemi et al. pioneered the use of geometric programming to recognize functions which were certifiably nonnegative by the am/gm inequality and a sums-of-binomial-squares representation [94, 95]. In 2012, Pantea, Koeppl, and Craciun derived an am/gm condition to certify $\mathbb{R}^n_+$-nonnegativity of polynomials supported on circuits [17, Theorem 3.6]. Follow-up work by August, Craciun, and Koeppl used [17, Theorem 3.6] to determine invariant sets of dynamical systems arising in biology [29]. A short while later, Iliman and de Wolff suggested taking sums of globally nonnegative circuit polynomials [96]. The resulting *SONC polynomials* have since become an established topic in the literature [97–101].

We continue to make connections to the am/gm-certificate literature throughout this chapter; [17, 96, 98, 99] are revisited in Subsection 3.5.3, and [96, 100, 101] are addressed in Subsection 3.5.4.

3.1.5 Notation

Our notation follows Chapter 2. The most important aspects of notation to keep in mind is our use of exponent vector sets $\mathcal{A} \subset \mathbb{R}^n$ as linear operators $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^{\mathcal{A}}$. Note that for $\boldsymbol{\nu} \in \mathbb{R}^{\mathcal{A}}$ and $\boldsymbol{\beta} \in \mathcal{A}$, the operation $\boldsymbol{\nu} \mapsto [\mathcal{A} \setminus \boldsymbol{\beta}]^{\dagger} \boldsymbol{\nu}_{\setminus \boldsymbol{\beta}}$ is well defined. If $\boldsymbol{\nu}$ belongs to $\mathbb{R}^{\mathcal{A}}$ and $\boldsymbol{\beta} \in \mathbb{R}^n$, then we evaluate $[\mathcal{A} - \boldsymbol{\beta}]^{\dagger} \boldsymbol{\nu}$ by identifying $\mathbb{R}^{\mathcal{A}} \cong \mathbb{R}^{[\mathcal{A} - \boldsymbol{\beta}]}$.

With the exception of the symbol C, capital letters in calligraphic font are treated as such hybrid point-sets and linear operators. General sets are given by capital letters in sans-serif font. This chapter refers to signomials by either writing them out in full or with the abbreviation $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$.

Given two sets S, T in a common vector space, we have the Minkowski sum $\mathsf{S} + \mathsf{T} = \{\boldsymbol{x} + \boldsymbol{y} \; : \; \boldsymbol{x} \in \mathsf{S}, \; \boldsymbol{y} \in \mathsf{T}\}$. We extend the exponential function $\exp : \mathbb{R} \to \mathbb{R}$ first to vectors in an elementwise fashion and then to sets in a pointwise fashion. So for a vector $\boldsymbol{x} \in \mathbb{R}^{\mathsf{S}}$ we have $\exp \boldsymbol{x} = (e^{x_s})_{s \in \mathsf{S}}$, and for a set $\mathsf{T} \subset \mathbb{R}^{\mathsf{S}}$ we have $\exp \mathsf{T} = \{\exp \boldsymbol{x} \; : \; \boldsymbol{x} \in \mathsf{T}\}$.

We often call a point set *simplicial* if it is affinely independent. The convex hull of a set T is denoted $\operatorname{conv} \mathsf{T}$, and the extreme points of a convex set T are given by $\operatorname{ext} \mathsf{T}$. A compact convex set is called simplicial if its extreme points are affinely independent. The relative entropy function $D : \mathbb{R}^{\mathsf{S}} \times \mathbb{R}^{\mathsf{S}} \to \mathbb{R} \cup \{+\infty\}$ continuously extends $D(\boldsymbol{u}, \boldsymbol{v}) = \sum_{s \in \mathsf{S}} u_s \log(u_s / v_s)$ to the product of nonnegative orthants $\mathbb{R}_+^{\mathsf{S}} \times \mathbb{R}_+^{\mathsf{S}}$.

3.2 Preliminaries on nonnegative signomials and signomial optimization

The cone of coefficients for nonnegative signomials over exponents $\mathcal{A}$ is

$$C_{\mathsf{NNS}}(\mathcal{A}) = \{\, \boldsymbol{c} \in \mathbb{R}^{\mathcal{A}} \; : \; \operatorname{Sig}(\mathcal{A}, \boldsymbol{c})(\boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{x} \text{ in } \mathbb{R}^n \,\}.$$

We sometimes overload terminology and refer to $C_{\mathsf{NNS}}(\mathcal{A})$ as a cone of signomials, rather than a cone of coefficients. Because $\mathcal{A}$ is a set (and hence contains no duplicates), the functions $\{\boldsymbol{x} \mapsto \exp\langle \boldsymbol{\alpha}, \boldsymbol{x}\rangle\}_{\boldsymbol{\alpha} \in \mathcal{A}}$ are linearly independent on $\mathbb{R}^n$ – this tells us that $C_{\mathsf{NNS}}(\mathcal{A})$ is pointed.

3.2.1 Primal and dual AGE cones

For an exponent $\boldsymbol{\beta} \in \mathcal{A}$, we define the $\boldsymbol{\beta}^{th}$ *AGE cone*

$$C_{\mathsf{AGE}}(\mathcal{A}, \boldsymbol{\beta}) = \{\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}} : \boldsymbol{c}_{\backslash \boldsymbol{\beta}} \geq \boldsymbol{0} \text{ and } \boldsymbol{c} \text{ belongs to } C_{\mathsf{NNS}}(\mathcal{A})\}. \qquad (3.1)$$

It is evident that $C_{\mathsf{AGE}}(\mathcal{A}, \boldsymbol{\beta})$ is a full-dimensional pointed convex cone which contains the nonnegative orthant. By using a convex duality argument, [13] shows that a vector $\boldsymbol{c}$ with $\boldsymbol{c}_{\backslash \boldsymbol{\beta}} \geq \boldsymbol{0}$ belongs to $C_{\mathsf{AGE}}(\mathcal{A}, \boldsymbol{\beta})$ if and only if some $\boldsymbol{\nu} \in \mathbb{R}^{\mathcal{A}}$ satisfies

$$\langle \boldsymbol{1}, \boldsymbol{\nu} \rangle = 0, \qquad \mathcal{A}^{\dagger} \boldsymbol{\nu} = \boldsymbol{0}, \quad \text{and} \quad D(\boldsymbol{\nu}_{\backslash \boldsymbol{\beta}}, e\boldsymbol{c}_{\backslash \boldsymbol{\beta}}) \leq c_{\boldsymbol{\beta}}. \qquad (3.2)$$

It is crucial that the representation in (3.2) is jointly convex in $\boldsymbol{c}$ and the auxiliary variable $\boldsymbol{\nu}$, and moreover that no assumption is made on the sign of $c_{\boldsymbol{\beta}}$.

Using the representation (3.2), one may derive the following expression for the dual of the $\boldsymbol{\beta}^{\text{th}}$ AGE cone

$$C_{\mathsf{AGE}}(\mathcal{A}, \boldsymbol{\beta})^{\dagger} = \operatorname{cl}\{\boldsymbol{\nu} \in \mathbb{R}^{\mathcal{A}} : \boldsymbol{\nu} > \boldsymbol{0}, \text{ and for some } \boldsymbol{\mu} \text{ in } \mathbb{R}^n \text{ we have}$$

$$\nu_{\boldsymbol{\beta}} \log(\nu_{\boldsymbol{\alpha}} / \nu_{\boldsymbol{\beta}}) \geq \langle \boldsymbol{\alpha} - \boldsymbol{\beta}, \boldsymbol{\mu} \rangle \text{ for } \boldsymbol{\alpha} \text{ in } \mathcal{A}\}. \qquad (3.3)$$

The "size" of a primal or dual AGE cone refers to the number of variables plus the number of constraints in the above representations, which is $O(m)$ assuming $n \leq m$.

3.2.2 Primal and dual SAGE cones

Cones of coefficients for SAGE signomials can be obtained by Minkowski sums

$$C_{\mathsf{SAGE}}(\mathcal{A}) := \sum_{\beta \in \mathcal{A}} C_{\mathsf{AGE}}(\mathcal{A}, \beta). \tag{3.4}$$

Standard calculations in conic duality yield the following expression for a dual SAGE cone

$$C_{\mathsf{SAGE}}(\mathcal{A})^\dagger = \bigcap_{\beta \in \mathcal{A}} C_{\mathsf{AGE}}(\mathcal{A}, \beta)^\dagger. \tag{3.5}$$

Equations 3.4 and 3.5 provide natural definitions, but they also contain redundancies.

Proposition 3.2.1. *[13, Section 2.4] Let* $\mathsf{P} = \mathrm{conv}(\mathcal{A})$. *If* α *is an extreme point of* P *and* $f = \mathrm{Sig}(\mathcal{A}, c)$ *is nonnegative, then* $c_\alpha \geq 0$. *Consequently, if* β *is an extreme point of* P *then* $C_{\mathsf{AGE}}(\mathcal{A}, \beta) = \mathbb{R}_+^{\mathcal{A}}$.

Proposition 3.2.1 is the most basic way Newton polytopes appear in the analysis of nonnegative signomials. In our context it means that so long as $\mathrm{ext}\,\mathrm{conv}(\mathcal{A}) \subsetneq \mathcal{A}$, we can take $C_{\mathsf{SAGE}}(\mathcal{A})$ as the Minkowski sum of AGE cones $C_{\mathsf{AGE}}(\mathcal{A}, \beta)$ for the β that are nonextremal in $\mathrm{conv}(\mathcal{A})$.

3.2.3 SAGE relaxations

We explained in Section 2.4 how nonnegativity cones and moment cones can be used to turn nonconvex optimization problems into convex cone programs. We now formalize this procedure for signomials and SAGE certificates. Throughout, we assume $\mathbf{0} \in \mathcal{A}$ and take $f = \mathrm{Sig}(\mathcal{A}, c)$.

For unconstrained minimization, we simply write $f - \gamma = \mathrm{Sig}(\mathcal{A}, c - \gamma \delta_0)$ to find

$$f^\star_{\mathbb{R}^n} = \sup\{\gamma \ : \ c - \gamma \delta_0 \text{ in } C_{\mathsf{NNS}}(\mathcal{A})\}.$$

We produce lower bounds on $f^\star_{\mathbb{R}^n}$ by replacing $C_{\mathsf{NNS}}(\mathcal{A})$ with the smaller cone $C_{\mathsf{SAGE}}(\mathcal{A})$. Specifically, we have a primal-dual pair

$$f^{\mathsf{SAGE}}_{\mathbb{R}^n} := \sup\{ \ \gamma \ : \ c - \gamma \delta_0 \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})\} \tag{3.6}$$

$$= \inf\{ \ \langle c, v \rangle \ : \ \langle \delta_0, v \rangle = 1, \ v \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})^\dagger\} \leq f^\star_{\mathbb{R}^n}. \tag{3.7}$$

Equality of the primal and dual values is justified by Theorem 2.6.1, since $C_{\mathsf{SAGE}}(\mathcal{A})$ is closed and pointed and $x \mapsto 1$ is trivially a nonzero SAGE function.

We make use of Lagrangians to handle constrained problems; this process is explained briefly in Subsection 3.4.4.

3.3   Structural results for SAGE certificates

This section presents two new geometric results and analytical characterizations on the SAGE cone. These results have applications to polynomial nonnegativity, as discussed later in Section 3.5. Statements of the theorems are provided below along with remarks on the theorems' significance. Proofs are deferred to later subsections.

3.3.1   Summary of structural results

Our first theorem shows that when checking if c belongs to $C_{\mathsf{SAGE}}(\mathcal{A})$, we can restrict the search space of SAGE decompositions to those exhibiting a very particular structure. It highlights the sparsity-preserving property of SAGE, and in so doing has significant implications for both the practicality of solving SAGE relaxations, and Section 3.5's development of SAGE polynomials.

Theorem 3.3.1. *If c is a vector in $C_{\mathsf{SAGE}}(\mathcal{A})$ with nonempty $\mathcal{N} \coloneqq \{\alpha : c_\alpha < 0\}$, then there exist vectors $\{c^{(\beta)} \in C_{\mathsf{AGE}}(\mathcal{A}, \beta)\}_{\beta \in \mathcal{N}}$ satisfying $c = \sum_{\beta \in \mathcal{N}} c^{(\beta)}$ and $c_\alpha^{(\beta)} = 0$ for all distinct $\beta, \alpha \in \mathcal{N}$.*

We can use Theorem 3.3.1 to define some parameterized AGE cones that will be of use to us in Section 3.4. Specifically, for $\mathcal{N} \subset \mathcal{A}$ and $\beta \in \mathcal{A}$, define

$$C_{\mathsf{AGE}}(\mathcal{A}, \beta, \mathcal{N}) = \{\, c \in C_{\mathsf{AGE}}(\mathcal{A}, \beta) \,:\, c_\alpha = 0 \text{ for all } \alpha \text{ in } \mathcal{N} \setminus \beta\}.$$

In terms of such sets we have the following corollary of Theorem 3.3.1.

Corollary 3.3.2. *Suppose $\mathbf{0} \in \mathcal{A}$. A signomial $f = \mathrm{Sig}(\mathcal{A}, c)$ has*

$$f_{\mathbb{R}^n}^{\mathsf{SAGE}} = \sup\{\gamma : c - \gamma \delta_0 \text{ in } \textstyle\sum_{\beta \in \mathcal{N} \cup \{0\}} C_{\mathsf{AGE}}(\mathcal{A}, \beta, \mathcal{N})\}$$

for both $\mathcal{N} = \{\alpha : c_\alpha < 0\}$ and $\mathcal{N} = \{\alpha : c_\alpha \leq 0\}$.

This corollary has two implications concerning practical algorithms for signomial optimization. First, it shows that for $k = |\{\alpha \in \mathcal{A} : c_\alpha < 0\}|$, computing $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ can easily be accomplished with a relative entropy program of size $O(k|\mathcal{A}|)$; this is a dramatic improvement over the naive implementation for computing $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$, which involves a relative entropy program of size $O(|\mathcal{A}|^2)$. Second, the improved conditioning resulting from restricting the search space in this way often makes the difference in whether existing solvers can handle SAGE relaxations of moderate size. This point is highlighted in recent experimental demonstrations of relative entropy relaxations; the authors of [102] discuss various preprocessing strategies to more quickly solve such optimization problems.

Our next theorem characterizes the extreme rays of the SAGE cone. To describe these extreme rays, we use a notion from matroid theory [103, 104]: a set of points $\mathsf{X} = \{x_i\}_{i=1}^{\ell}$ is called a *circuit* if it is affinely dependent, but any proper subset $\{x_i\}_{i \neq k}$ is affinely independent. If the convex hull of a circuit with ℓ elements contains $\ell - 1$ extreme points, then we say the circuit is *simplicial*.

Theorem 3.3.3. *If $c \in \mathbb{R}^{\mathcal{A}}$ generates an extreme ray of $C_{\mathsf{SAGE}}(\mathcal{A})$, then $\operatorname{supp} c$ is either a singleton or a simplicial circuit.*

Theorem 3.3.3 can be viewed as a signomial generalization of a result by Reznick concerning *agiforms* [14, Theorem 7.1]. The theorem admits a partial converse: if $\mathcal{A}' \cup \{\beta\} \subset \mathcal{A}$ is a simplicial circuit with nonextremal term β, then there is an extreme ray of $C_{\mathsf{AGE}}(\mathcal{A}', \beta)$ supported on $\mathcal{A}' \cup \{\beta\}$. When specialized to the context of polynomials, this result gives us an equivalence between SAGE polynomials (suitably defined in Section 3.5) and the previously defined SONC polynomials [96], thus providing an efficient description of the latter set which was not known to be tractable.

3.3.2 Proof of the restriction theorem for SAGE decompositions (Theorem 3.3.1)

Our proof requires two lemmas. The first such lemma indicates the claim of the theorem applies far more broadly than for SAGE functions alone.

Lemma 3.3.4. *Let $\mathsf{K} \subset \mathbb{R}^m$ be a convex cone containing the nonnegative orthant. For an index $i \in [m]$, define $\mathsf{C}_i = \{c \in \mathsf{K} : c_{\backslash i} \geq 0\}$, and sum these to $\mathsf{C} = \sum_{i=1}^m \mathsf{C}_i$. We claim that a vector c with at least one negative entry belongs to C if and only if*

$$c \in \sum_{i : c_i < 0} \mathsf{C}_i.$$

Proof. Suppose $c \in \mathsf{C}$ has a decomposition $c = \sum_{i \in \mathsf{N}} c^{(i)}$ where each $c^{(i)}$ belongs to C_i. If $\mathsf{N} = \{i : c_i < 0\}$, then there is nothing to prove, so suppose there is some k in N with $c_k \geq 0$. We construct an alternative decomposition of c using only cones C_i with i in $\mathsf{N} \setminus \{k\}$.

The construction depends on the sign of $c_k^{(k)}$. If $c_k^{(k)}$ is nonnegative then the problem of removing dependence on C_k simple: for i in $\mathsf{N} \setminus \{k\}$, the vectors

$$\tilde{c}^{(i)} = c^{(i)} + c^{(k)}/(|N| - 1)$$

belong to C_i (since $C_i \supset \mathbb{R}^m_+$), and sum to c. If instead $c_k^{(k)} < 0$, then there exists some index $i \neq k$ in N with $c_k^{(i)}$ positive. This allows us to define the distribution λ with $\lambda_i = c_k^{(i)} / \sum_{j \in N \setminus \{k\}} c_k^{(j)}$ for $i \neq k$ in N. With λ we construct the $|N| - 1$ vectors

$$\tilde{c}^{(i)} = c^{(i)} + \lambda_i c^{(k)}.$$

The vectors $\tilde{c}^{(i)}$ belong to K because they are a conic combination of vectors in K ($c^{(i)}$ and $c^{(k)}$). We claim that for every $i \neq k$ in N, the coordinate $\tilde{c}_k^{(i)}$ is nonnegative. This is certainly true when $\lambda_i = 0$, but more importantly, $\lambda_i > 0$ implies

$$\frac{1}{\lambda_i} \tilde{c}_k^{(i)} = \frac{1}{\lambda_i} \left(c_k^{(i)} + \lambda_i c_k^{(k)} \right) = [\textstyle\sum_{j \in N \setminus \{k\}} c_k^{(j)}] + c_k^{(k)} = c_k \geq 0.$$

Hence c can be expressed as the sum of vectors $\{\tilde{c}^{(i)}\}_{i \in N \setminus \{k\}}$ where each vector $\tilde{c}^{(i)}$ belongs to C_i.

From here, update $N \leftarrow N \setminus \{k\}$. If N contains another index k' with $c_{k'} \geq 0$, then repeat the above procedure to remove the unnecessary cone $C_{k'}$. Naturally, this process continues until $N = \{i : c_i < 0\}$. $\qquad\square$

Lemma 3.3.5. *Let w, v be vectors in $\mathbb{R}^m$ with distinguished indices $i \neq j$ so that*

$$w_{\setminus i}, \; v_{\setminus j} \geq \mathbf{0} \quad \text{and} \quad w_k + v_k < 0 \text{ for } k \text{ in } \{i, j\}.$$

Then there exist vectors $\hat{w}, \hat{v}$ in the conic hull of $\{w, v\}$ which satisfy

$$\hat{w} + \hat{v} = w + v \quad \text{and} \quad \hat{w}_j = \hat{v}_i = 0.$$

Proof. By reindexing, take $i = 1$ and $j = 2$. We will decide $\lambda \in \mathbb{R}^4_+$ so that $\hat{w} = \lambda_2 w + \lambda_4 v$ and $\hat{v} = \lambda_1 w + \lambda_3 v$ satisfy the desired relations. One may verify that it is sufficient for λ in $\mathbb{R}^4_+$ to solve

$$\begin{bmatrix} w_1 & 0 & v_1 & 0 \\ 0 & w_2 & 0 & v_2 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}. \tag{3.8}$$

The determinant of the matrix above is $d = w_1 v_2 - v_1 w_2$. If w_2 or $v_1 = 0$, then $d > 0$. If $w_2, v_1 \neq 0$, then $d > 0 \Leftrightarrow |v_2/w_2| \cdot |w_1/v_1| > 1$. In this case we use the assumptions on w, v to establish the slightly stronger condition that $|v_2/w_2| > 1$ and $|w_1/v_1| > 1$. In both cases we have a nonzero determinant, so there exists a unique λ in $\mathbb{R}^4$ satisfying system (3.8). Now we need only prove that this λ is nonnegative.

One may verify that the symbolic solution to (3.8) is

$$\lambda_1 = -(w_2 + v_2)v_1/d, \quad \lambda_2 = (w_1 + v_1)v_2/d,$$
$$\lambda_3 = w_1(w_2 + v_2)/d, \quad \lambda_4 = -(w_1 + v_1)w_2/d,$$

and furthermore that all numerators and denominators are nonnegative.   □

Theorem 3.3.1. In this proof we consider $\mathcal{A} \in \mathbb{R}^{m \times n}$ as built from rows α_i, we identify $\mathbb{R}^{\mathcal{A}} = \mathbb{R}^m$, and we use $C_{\mathsf{AGE}}(\mathcal{A}, i) = C_{\mathsf{AGE}}(\mathcal{A}, \alpha_i)$. Let $c^\star$ be a vector in $C_{\mathsf{SAGE}}(\mathcal{A})$ with k negative entries $c_1^\star, \ldots, c_k^\star$. It is clear that the AGE cones $C_{\mathsf{AGE}}(\mathcal{A}, i)$ satisfy the hypothesis of Lemma 3.3.4, with $\mathsf{K} = C_{\mathsf{NNS}}(\mathcal{A})$. Therefore there exists a k-by-m matrix C with i^{th} row $c_i \in C_{\mathsf{AGE}}(\mathcal{A}, i)$, and $c^\star = \sum_{i=1}^{k} c_i$. We prove the result by transforming C into a matrix with rows c_i satisfying the required properties, using only row-sum preserving conic combinations from Lemma 3.3.5.

It is clear that for any pair of distinct i, j, the vectors c_i, c_j satisfy the hypothesis of Lemma 3.3.5, thus there exist $\hat{c}_i, \hat{c}_j$ in the conic hull of c_i, c_j where $\hat{c}_{ij} = \hat{c}_{ji} = 0$ and $\hat{c}_i + \hat{c}_j = c_i + c_j$. Furthermore, this remains true if we modify C by *replacing* $(c_i, c_j) \leftarrow (\hat{c}_i, \hat{c}_j)$.

We proceed algorithmically: apply Lemma 3.3.5 to rows $(1, 2)$, then $(1, 3)$, and continuing to rows $(1, k)$. At each step of this process we eliminate $c_{j1} = 0$ for $j > 1$ and maintain $c_{ji} \geq 0$ for off-diagonal c_{ji}. We then apply the procedure to the second column of C, beginning with rows 2 and 3. Since $c_{j1} = 0$ for $j > 1$, none of the row operations introduce an additional nonzero in the first column of C, and so the first column remains zero below c_{11}, and the second column becomes zero below c_{22}. Following this pattern we reduce C to have zeros on the strictly lower-triangular block in the first k columns, in particular terminating with $c_{kk} = c_k^\star < 0$.

The next phase is akin to back-substitution. Apply Lemma 3.3.5 to rows $(k, k-1)$, then $(k, k-2)$, and continue until rows $(k, 1)$. This process zeros out the kth column of C above c_{kk}. The same procedure applies with rows $(k-1, k-2)$, then $(k-1, k-3)$, through $(k-1, 1)$, to zero the $(k-1)^{\text{st}}$ column of C except for the single entry $c_{[k-1][k-1]} = c_{k-1}^\star < 0$. The end result of this process is that the first k columns of C comprise a diagonal matrix with entries $(c_1^\star, \ldots, c_k^\star) < \mathbf{0}$.

The resulting matrix C satisfies the claimed sparsity conditions. Since all row-operations involved conic combinations, each row of the resulting matrix C defines a nonnegative signomial. The theorem follows since row i of the resulting matrix has a single negative component $c_{ii} = c_i^\star < 0$.   □

3.3.3 Proof of extreme ray characterization of the SAGE cone (Theorem 3.3.3)

Because every ray in the SAGE cone (extreme or otherwise) can be written as a sum of rays in AGE cones, it suffices to characterize the extreme rays of AGE cones. For the duration of this section we discuss the AGE cone $C_{\mathsf{AGE}}(\mathcal{A}, \beta)$, where β is nonextremal in $\mathrm{conv}(\mathcal{A})$.

It can easily be shown that for any index β in $\mathcal{A}$, the ray $\{r\delta_\beta : r \geq 0\}$ is extremal in $C_{\mathsf{AGE}}(\mathcal{A}, \beta)$. We call these rays (those supported on a single coordinate) the *atomic* extreme rays of the AGE cone. The work in showing Theorem 3.3.3 is to prove that all nonatomic extreme rays of the AGE cone are supported on simplicial circuits. Our proof will appeal to the following basic fact concerning polyhedral geometry, which we establish in the appendix.

Lemma 3.3.6. *Fix $B \in \mathbb{R}^{n \times d}$, $h \in \mathbb{R}^n$, and $\Lambda = \{\lambda \in \Delta_d : B\lambda = h\}$. For any $\lambda \in \Lambda$, there exist $\{\lambda^{(i)}\}_{i=1}^{\ell} \subset \Lambda$ and $\theta \in \Delta_\ell$ for which $\{b_j : \lambda_j^{(i)} > 0\}$ are affinely independent, and $\lambda = \sum_{i \in [\ell]} \theta_i \lambda^{(i)}$.*

Theorem 3.3.3. Let $\mathsf{N}_\beta = \{v \in \mathbb{R}^{\mathcal{A}} : v_{\backslash\beta} \geq 0, \langle 1, v \rangle = 0\}$. We seek an $\ell \in \mathbb{N}$ where we can decompose $c \in C_{\mathsf{AGE}}(\mathcal{A}, \beta)$ as a sum of $\ell+1$ AGE vectors $\{c^{(i)}\}_{i=1}^{\ell+1} \subset C_{\mathsf{AGE}}(\mathcal{A}, \beta)$, where $\mathrm{supp}\, c^{(i)}$ are simplicial circuits for $i \in [\ell]$ and $c^{(\ell+1)} \geq 0$. Since c is an AGE vector, there is an associated $v \in \mathsf{N}_\beta$ for which $\mathcal{A}^\dagger v = 0$ and $D(v_{\backslash\beta}, ec_{\backslash\beta}) \leq c_\beta$. If v is zero, then $D(v_{\backslash\beta}, ec_{\backslash\beta}) = 0 \leq c_\beta$, so $\ell = 0$ and $c^{(\ell+1)} = c$ provides the required decomposition. The interesting case, of course, is when v is nonzero. We proceed by providing a mechanism to decompose v into a convex combination of certain vectors $\{v^{(i)}\}_{i=1}^{\ell}$, and from there we obtain suitable AGE vectors $c^{(i)}$ from each $v^{(i)}$.

Given $v \neq 0$, the vector $\lambda := v_{\backslash\beta}/|v_\beta|$ belongs to the probability simplex $\Delta_{\mathcal{A}\backslash\beta}$. We introduce this λ because $\mathcal{A}^\dagger v = 0$ is equivalent to $[\mathcal{A} \backslash \beta]^\dagger \lambda = \beta$, and the latter form is amenable to Lemma 3.3.6. Apply Lemma 3.3.6 to decompose λ into a convex combination of vectors $\{\lambda^{(i)}\}_{i=1}^{\ell} \subset \Delta_{\mathcal{A}\backslash\beta}$ for which $\mathrm{supp}\, \lambda^{(i)}$ are simplicial and $\lambda^{(i)}$ satisfy $[\mathcal{A} \backslash \beta]^\dagger \lambda^{(i)} = \beta$; let $\theta \in \Delta_\ell$ denote the vector of convex combination coefficients for this decomposition of λ. For each $\lambda^{(i)}$, define $v^{(i)}$ by $v_{\backslash\beta}^{(i)} = \lambda^{(i)}|v_\beta|$ and $v_\beta^{(i)} = v_\beta$. These values for $v^{(i)}$ evidently satisfy $\mathcal{A}^\dagger v^{(i)} = 0$ and $\sum_{i \in [\ell]} \theta_i v^{(i)} = v$. From these $v^{(i)}$ we construct $c^{(i)} \in \mathbb{R}^{\mathcal{A}}$ by

$$c_\alpha^{(i)} = \begin{cases} (c_\alpha/v_\alpha)v_\alpha^{(i)} & \text{if } v_\alpha > 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \alpha \neq \beta,$$

and for $\alpha = \beta$ we take $c_\alpha^{(i)} = D(v_{\backslash \beta}^{(i)}, ec_{\backslash \beta}^{(i)})$.

By construction these $c^{(i)}$ belong to $C_{\mathsf{AGE}}(\mathcal{A}, \beta)$, and supp $c^{(i)}$ are simplicial circuits. We now take a componentwise approach to showing $\sum_{i \in [\ell]} \theta_i c^{(i)} \leq c$. For indices $\alpha \neq \beta$ with $v_\alpha > 0$, the inequality actually holds with *equality*

$$\sum_{i \in [\ell]} \theta_i c_\alpha^{(i)} = (c_\alpha/v_\alpha)(\textstyle\sum_{i \in [\ell]} \theta_i v_\alpha^{(i)}) = c_\alpha.$$

Now we turn to showing $\sum_{i \in [\ell]} \theta_i c_\beta^{(i)} \leq c_\beta$; we specifically claim that

$$\sum_{i \in [\ell]} \theta_i c_\beta^{(i)} = \sum_{i \in [\ell]} \theta_i D(v_{\backslash \beta}^{(i)}, ec_{\backslash \beta}^{(i)}) = D(v_{\backslash \beta}, ec_{\backslash \beta}) \leq c_\beta. \tag{3.9}$$

For the three relations in display (3.9), the first holds from the definitions of $c_\beta^{(i)}$, and the last holds from our assumptions on (c, v), so only the second equality needs explaining. For this we use the fact that definitions of $c_\alpha^{(i)}$ relative to $v_\alpha^{(i)}$ preserve ratios with c_α relative to v_α, i.e.

$$D(v_{\backslash \beta}^{(i)}, ec_{\backslash \beta}^{(i)}) = \sum_{\alpha \neq \beta} v_\alpha^{(i)} \log\left(\frac{v_\alpha^{(i)}}{ec_\alpha^{(i)}}\right) = \sum_{\alpha \neq \beta} v_\alpha^{(i)} \log\left(\frac{v_\alpha}{ec_\alpha}\right). \tag{3.10}$$

One may then prove the middle equality in display (3.9) by summing $\theta_i D(v_{\backslash \beta}^{(i)}, c_{\backslash \beta})$ over i, applying the identity in equation (3.10), and then interchanging the sums over i and α. Formally,

$$\sum_{i \in [\ell]} \theta_i D(v_{\backslash \beta}^{(i)}, ec_{\backslash \beta}^{(i)}) = \sum_{\alpha \neq \beta} \log\left(\frac{v_\alpha}{ec_\alpha}\right) \overbrace{\left(\sum_{i \in [\ell]} \theta_i v_\alpha^{(i)}\right)}^{=v_\alpha} = D(v_{\backslash \beta}, ec_{\backslash \beta}).$$

We have effectively established the claim of the theorem. To find a decomposition of the form desired at the beginning of this proof, one rescales $c^{(i)} \leftarrow \theta_i c^{(i)}$ and sets $c^{(\ell+1)} = c - \sum_{i \in [\ell]} c^{(i)}$. $\qquad \square$

3.4 The role of Newton polytopes in SAGE signomials

This section begins by introducing two theorems (Theorems 3.4.1 and 3.4.2) concerning SAGE representability versus signomial nonnegativity. These theorems are then combined to obtain a third theorem (Theorem 3.4.3), which provides the most general yet-known conditions for when the SAGE and nonnegativity cones coincide. The proofs of Theorems 3.4.1 and 3.4.2 are contained in Subsections 3.4.2

and 3.4.3. Applications of Theorem 3.4.1 are given in Subsection 3.4.4. Subsection 3.4.5 uses a distinct proof strategy (nevertheless Newton-polytope based) to determine a condition on when SAGE can recognize signomials which are bounded below.

3.4.1 When SAGE recovers the nonnegativity cone

The following theorem is the first instance beyond AGE functions when SAGE-representability is known to be equivalent to nonnegativity.

Theorem 3.4.1. *Suppose* $\operatorname{ext conv}(\mathcal{A})$ *is simplicial, and that c has $c_{\alpha} \leq 0$ whenever α is nonextremal. Then c belongs to $C_{\mathsf{SAGE}}(\mathcal{A})$ if and only if c belongs to $C_{\mathsf{NNS}}(\mathcal{A})$.*

Our proof of the theorem (Subsection 3.4.2) uses convex duality in a central way, and provides intuition for why the theorem's assumptions are needed. Section 3.6 provides counter-examples to relaxations of Theorem 3.4.1 obtained through weaker hypothesis.

This section's next theorem (proven in Subsection 3.4.3) concerns conditions on $\mathcal{A}$ for when the SAGE and nonnegativity cones can be expressed as a Cartesian product of simpler sets. To aid in exposition we introduce a definition: a set $\mathcal{A}$ can be *partitioned into k faces* if it can be written as a disjoint union $\mathcal{A} = \cup_{i \in [k]} \mathcal{A}^{(i)}$ where $\operatorname{conv}(\mathcal{A}^{(i)})$ are mutually disjoint faces of $\operatorname{conv}(\mathcal{A})$.
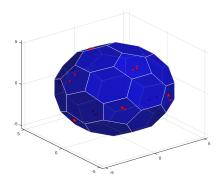
Theorem 3.4.2. *If $\{\mathcal{A}^{(i)}\}_{i=1}^{k}$ partition $\mathcal{A}$ into faces, then*
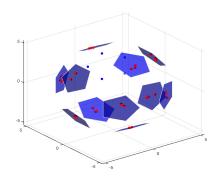
$$C_{\mathsf{NNS}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A}^{(1)}) \times \cdots \times C_{\mathsf{NNS}}(\mathcal{A}^{(k)})$$

and

$$C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{SAGE}}(\mathcal{A}^{(1)}) \times \cdots \times C_{\mathsf{SAGE}}(\mathcal{A}^{(k)}).$$

The following figure illustrates partitioning a set $\mathcal{A}$ where $\operatorname{ext conv}(\mathcal{A})$ are the vertices of the truncated icosahedron, and nonextremal terms (marked in red) lay in the relative interiors of certain pentagonal faces.

Note that every set $\mathcal{A}$ admits the trivial partition with $k = 1$. In fact, a natural regularity condition (one that we consider in Section 3.6) would be that $\mathcal{A}$ *only* admits the trivial partition. Regularity conditions aside, Theorems 3.4.1 and 3.4.2 can be combined with known properties of AGE functions to establish new conditions for when the SAGE and nonnegativity cones coincide.

Theorem 3.4.3. *Suppose $\mathcal{A}$ can be partitioned into faces where (1) simplicial faces contain at most two nonextremal exponents, and (2) all other faces contain at most one nonextremal exponent. Then $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$.*

Proof. Let $\mathcal{A}$ satisfy the assumptions of Theorem 3.4.3 with associated faces $\{\mathsf{F}_i\}_{i=1}^k$ and subsets $\mathcal{A}^{(i)}$, and fix c in $C_{\mathsf{NNS}}(\mathcal{A})$. For i in $[k]$, define the vector $c^{(i)}$ so that $c = \oplus_{i=1}^k c^{(i)}$ is the vector concatenation of the $c^{(i)}$. By Theorem 3.4.2, the condition $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$ holds if and only if $C_{\mathsf{SAGE}}(\mathcal{A}^{(i)}) = C_{\mathsf{NNS}}(\mathcal{A}^{(i)})$ for all i in $[k]$. Because we assumed that c belongs to $C_{\mathsf{NNS}}(\mathcal{A})$ it suffices to show that each $c^{(i)}$ belongs to $C_{\mathsf{SAGE}}(\mathcal{A}^{(i)})$.

Per Proposition 3.2.1, any vector $c^{(i)} \in C_{\mathsf{NNS}}(\mathcal{A}^{(i)})$ cannot have a negative entry $c_{\alpha}^{(i)}$ when α is extremal in $\mathrm{conv}(\mathcal{A}^{(i)})$. By assumption, $\mathcal{A}^{(i)}$ has at most two nonextremal terms, and so $c^{(i)} \in C_{\mathsf{NNS}}(\mathcal{A}^{(i)})$ can have at most two negative entries. If $c^{(i)}$ has at most one negative entry, then $c^{(i)}$ is an AGE vector. If on the other hand $c^{(i)}$ has two negative entries $c_{\alpha}^{(i)}, c_{\beta}^{(i)}$, then both of these entries must correspond to nonextremal α, β, and F_i must be simplicial. This allows us to invoke Theorem 3.4.1 on $c^{(i)}$ to conclude $c^{(i)} \in C_{\mathsf{SAGE}}(\mathcal{A}^{(i)})$. The result follows. $\square$

3.4.2 Simplicial sign patterns for SAGE versus nonnegativity (Theorem 3.4.1)

The proof of Theorem 3.4.1 begins by exploiting two key facts about signomials and SAGE relaxations: (1) that $C_{\mathsf{SAGE}}(\mathcal{A})$ and $C_{\mathsf{NNS}}(\mathcal{A})$ are invariant under translation of the exponent set $\mathcal{A}$, and (2) that strong duality *always* holds when computing

$f_{\mathbb{R}^n}^{\mathsf{SAGE}}$. These properties allow us to reduce the problem of checking SAGE decomposability to the problem of exactness of a convex relaxation for a signomial optimization problem.

Theorem 3.4.1. In this proof we take $\mathcal{A} = \{\boldsymbol{\alpha}_i\}_{i=1}^m \subset \mathbb{R}^n$. Begin by translating $\mathcal{A}$ to $\mathcal{A} \leftarrow \mathcal{A} - \boldsymbol{\alpha}_i$ where $\boldsymbol{\alpha}_i$ is an arbitrary extremal element of $\mathrm{conv}(\mathcal{A})$. Next, re-index the $\boldsymbol{\alpha}_i$ so that $\boldsymbol{\alpha}_1 = \mathbf{0}$. Fix $\boldsymbol{c}$ in $C_{\mathsf{NNS}}(\mathcal{A})$ and define $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ so that $f_{\mathbb{R}^n}^\star \geq 0$. We show that $f_{\mathbb{R}^n}^{\mathsf{SAGE}} = f_{\mathbb{R}^n}^\star$, thereby establishing $\boldsymbol{c} \in C_{\mathsf{SAGE}}(\mathcal{A})$.

Let $\mathsf{N} = \{i : c_i \leq 0\}$ and $\mathsf{E} = [m] \setminus \mathsf{N}$. Apply Corollary 3.3.2 to obtain a primal SAGE relaxation with dimension reduction, and then dualize that relaxation. By Theorem 2.6.1 (concerning strong duality) and Proposition 2.6.4 (concerning representations of dual cones contained in the nonnegative orthant), the SAGE bound can be expressed as

$$f_{\mathbb{R}^n}^{\mathsf{SAGE}} = \inf \ \langle \boldsymbol{c}, \boldsymbol{v} \rangle \tag{3.11}$$
$$\text{s.t. } \boldsymbol{v} \text{ in } \mathbb{R}_{++}^m \text{ has } v_1 = 1, \text{ and there exist } \{\boldsymbol{\mu}_i\}_{i \in \mathsf{N} \cup \{1\}} \subset \mathbb{R}^n \text{ with}$$
$$v_i \log(v_i/v_j) \leq \langle \boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j, \boldsymbol{\mu}_i \rangle \text{ for } j \text{ in } \mathsf{E} \text{ and } i \text{ in } \mathsf{N} \cup \{1\}.$$

In order to show $f_{\mathbb{R}^n}^{\mathsf{SAGE}} = f_{\mathbb{R}^n}^\star$, we reformulate (3.11) as the problem of computing $f_{\mathbb{R}^n}^\star$ by appropriate changes of variables and constraints.

We begin with a change of constraints. By the assumption that $c_i \leq 0$ for all nonextremal $\boldsymbol{\alpha}_i$, the set E satisfies $\{\boldsymbol{\alpha}_i\}_{i \in E} \subset \mathrm{ext\,conv}(\mathcal{A})$. Combine this with extremality of $\mathbf{0} = \boldsymbol{\alpha}_1$ and the assumption that $\mathrm{ext\,conv}(\mathcal{A})$ is simplicial to conclude that $\{\boldsymbol{\alpha}_i : i \text{ in } \mathsf{E} \setminus \{1\}\}$ are linearly independent. The linear independence of these vectors ensures that for fixed $\boldsymbol{v}$ we can always choose $\boldsymbol{\mu}_1$ to satisfy the following constraints *with equality*

$$v_1 \log(v_1/v_j) \leq \langle \boldsymbol{\alpha}_1 - \boldsymbol{\alpha}_j, \boldsymbol{\mu}_1 \rangle \text{ for all } j \text{ in } \mathsf{E}.$$

Therefore we can equivalently reformulate $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ as

$$f_{\mathbb{R}^n}^{\mathsf{SAGE}} = \inf \ \langle \boldsymbol{c}, \boldsymbol{v} \rangle$$
$$\text{s.t. } \boldsymbol{v} \text{ in } \mathbb{R}_{++}^m \text{ has } v_1 = 1, \text{ and there exist } \{\boldsymbol{\mu}_i\}_{i \in \mathsf{N} \cup \{1\}} \subset \mathbb{R}^n$$
$$\text{with } \log(v_j) = \langle \boldsymbol{\alpha}_j, \boldsymbol{\mu}_1 \rangle \text{ for all } j \text{ in } \mathsf{E}, \text{ and}$$
$$v_i \log(v_i/v_j) \leq \langle \boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j, \boldsymbol{\mu}_i \rangle \text{ for } j \text{ in } \mathsf{E}, \ i \text{ in } \mathsf{N}.$$

Next we rewrite the constraint $v_i \log(v_i/v_j) \leq \langle \boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j, \boldsymbol{\mu}_i \rangle$ as $\log(v_i) - \log(v_j) \leq \langle \boldsymbol{\alpha}_i - \boldsymbol{\alpha}_j, \boldsymbol{\mu}_i \rangle$ by absorbing v_i into $\boldsymbol{\mu}_i$. If we also substitute the expression for $\log(v_j)$

given by the equality constraints, then the inequality constraints become

$$\log(v_i) \leq \langle \alpha_i, \mu_i \rangle + \langle \alpha_j, \mu_1 - \mu_i \rangle \text{ for all } j \text{ in } \mathsf{E}, \; i \text{ in } \mathsf{N}. \tag{3.12}$$

We now show that for every i in N, the choice $\mu_i = \mu_1$ makes these inequality constraints as loose as possible.

Towards this end, define $\psi_i(x) = \langle \alpha_i, x \rangle + \min_{j \in E}\{\langle \alpha_j, (\mu_1 - x) \rangle\}$; note that for fixed i and μ_i, the number $\psi_i(\mu_i)$ is the minimum over all $|E|$ right hand sides in (3.12). It is easy to verify that ψ_i is concave, and because of this we know that ψ_i is maximized at $x^\star$ if and only if $0 \in (\partial \psi_i)(x^\star)$. Standard subgradient calculus tells us that $(\partial \psi_i)(x)$ is precisely the convex hull of vectors $\alpha_i - \alpha_k$ where k is an index at which the minimum (over $j \in E$) is obtained. Therefore $(\partial \psi_i)(\mu_1) = \mathrm{conv}\{\alpha_i - \alpha_j : j \text{ in } \mathsf{E}\}$, and this set must contain the zero vector (unless perhaps $c_i = 0$, in which case the constraints on v_i are inconsequential). Hence $\max_{x \in \mathbb{R}^n}\{\psi_i(x)\} = \langle \alpha_i, \mu_1 \rangle$, and so inequality constraints (3.12) reduce to

$$\log(v_i) \leq \langle \alpha_i, \mu_1 \rangle \text{ for all } i \text{ in } \mathsf{N}. \tag{3.13}$$

Since the objective $\langle c, v \rangle$ is decreasing in v_i for i in N, we can actually take the constraints in (3.13) to be binding. We established much earlier that $v_i = \exp\langle \alpha_i, \mu_1 \rangle$ for i in E. Taking these together we see $v_i = \exp\langle \alpha_i, \mu_1 \rangle$ for all i, and so

$$f_{\mathbb{R}^n}^{\mathsf{SAGE}} = \inf\{ \; \textstyle\sum_{i=1}^m c_i \exp\langle \alpha_i, \mu_1 \rangle \; : \; \mu_1 \text{ in } \mathbb{R}^n\} = f_{\mathbb{R}^n}^\star \tag{3.14}$$

as required. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Let us now recap how the assumptions of Theorem 3.4.1 were used at various stages in the proof. For one thing, all discussion up to and including the statement of Problem (3.11) was fully general; the expression for $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ used none of the assumptions of the theorem. The next step was to use linear independence of nonzero extreme points to allow us to satisfy $v_1 \log(v_1/v_j) \leq \langle \alpha_1 - \alpha_j, \mu_1 \rangle$ with equality. The reader can verify that if we did not have linear independence, but we *were* told that those constraints were binding at the optimal $v^\star$, then we would still have $f_{\mathbb{R}^n}^{\mathsf{SAGE}} = f_{\mathbb{R}^n}^\star$ under the stated sign pattern assumption on c. Note how the sign pattern assumption on c was only really used to replace $\log(v_i) \leq \langle \alpha_i, \mu_1 \rangle$ from (3.13) by $\log(v_i) = \langle \alpha_i, \mu_i \rangle$.

3.4.3 Proof of the partitioning theorem (Theorem 3.4.2)

The following lemma adapts claim (iv) from Theorem 3.6 of Reznick [14] to signomials. Because the lemma is important for our subsequent theorems, the appendix contains a more complete proof than can be found in Reznick's [14]. As a matter of notation: for any face F of conv $\mathcal{A}$, write $\mathrm{Sig}_\mathsf{F}(\mathcal{A}, \boldsymbol{c})$ to mean the signomial with exponents $\boldsymbol{\alpha} \in \mathsf{F} \cap \mathcal{A}$ and corresponding coefficients $c_{\boldsymbol{\alpha}}$.

Lemma 3.4.4. *If* F *is a face of the polytope* conv$(\mathcal{A})$ *then* $\mathrm{Sig}_\mathsf{F}(\mathcal{A}, \boldsymbol{c})^\star_{\mathbb{R}^n} < 0$ *implies* $\mathrm{Sig}(\mathcal{A}, \boldsymbol{c})^\star_{\mathbb{R}^n} < 0$.

Theorem 3.4.2. Let $\mathcal{A}$ have partition $\mathcal{A} = \cup_{i \in [k]} \mathcal{A}^{(i)}$. It is clear from the definition of the SAGE cone that $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{SAGE}}(\mathcal{A}^{(1)}) \times \cdots C_{\mathsf{SAGE}}(\mathcal{A}^{(k)})$. The bulk of this proof is to show that $C_{\mathsf{NNS}}(\mathcal{A})$ admits the same decomposition.

Let $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ for some $\boldsymbol{c}$ in $\mathbb{R}^\mathcal{A}$. The vector $\boldsymbol{c}$ is naturally decomposed into a concatenation $\boldsymbol{c} = \boldsymbol{c}^{(1)} \oplus \cdots \oplus \boldsymbol{c}^{(k)}$ of smaller vectors $\boldsymbol{c}^{(i)} \in \mathbb{R}^{\mathcal{A}^{(i)}}$. For each i in $[k]$ define $f^{(i)} = \mathrm{Sig}(\mathcal{A}^{(i)}, \boldsymbol{c}^{(i)})$ so that $f = \sum_{i=1}^k f^{(i)}$. If any $(f^{(i)})^\star_{\mathbb{R}^n}$ is negative, then Lemma 3.4.4 tells us that $f^\star_{\mathbb{R}^n}$ must also be negative. Meanwhile if all $(f^{(i)})^\star_{\mathbb{R}^n}$ are nonnegative, then the same must be true of $f^\star_{\mathbb{R}^n} \geq \sum_{i=1}^k (f^{(i)})^\star_{\mathbb{R}^n}$. The result follows. $\qquad\square$

3.4.4 Corollaries for signomial programming

Signomial minimization is naturally related via duality to checking signomial nonnegativity. Thus we build on groundwork laid in Sections 3.3 and 3.4 to obtain consequences for signomial minimization.

Corollary 3.4.5. *Assume* conv$(\mathcal{A})$ *is simplicial, that* $\boldsymbol{0} \in \mathcal{A}$, *and that nonzero nonextremal* $\boldsymbol{\alpha}$ *have* $c_{\boldsymbol{\alpha}} \leq 0$. *Then either* $f^{\mathsf{SAGE}}_{\mathbb{R}^n} = f^\star_{\mathbb{R}^n}$, *or* $f^\star_{\mathbb{R}^n} \in (f^{\mathsf{SAGE}}_{\mathbb{R}^n}, c_{\boldsymbol{0}})$.

Proof. It suffices to show that $f^{\mathsf{SAGE}}_{\mathbb{R}^n} < f^\star_{\mathbb{R}^n}$ implies $f^\star_{\mathbb{R}^n} < c_{\boldsymbol{0}}$. This follows as the contrapositive of the following statement: "If $f^\star_{\mathbb{R}^n} \geq c_{\boldsymbol{0}}$, then by Theorem 3.4.1 the nonnegative signomial $f - f^\star_{\mathbb{R}^n}$ is SAGE, which in turn ensures $f^{\mathsf{SAGE}}_{\mathbb{R}^n} = f^\star_{\mathbb{R}^n}$." $\qquad\square$

Now we consider constrained signomial programs. As above, we assume $\mathcal{A}$ contains the zero vector. Starting with problem data (f, g) where $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ and $g_j = \mathrm{Sig}(\mathcal{A}, \boldsymbol{g}_j)$ for j in $[k]$, consider the problem of computing

$$(f, g)^\star := \inf\{f(\boldsymbol{x}) \ : \ \boldsymbol{x} \text{ in } \mathbb{R}^n \text{ satisfies } g(\boldsymbol{x}) \geq \boldsymbol{0}\}. \tag{3.15}$$

It is evident[1] that we can relax the problem to that of

$$(f, g)_{\mathsf{SAGE}} := \inf\{ \langle c, v \rangle : v \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})^\dagger$$
$$\text{satisfies } v_0 = 1 \text{ and } G^\dagger v \geq 0\} \leq (f, g)^\star$$

where G is the $m \times k$ matrix whose columns are the g_j.

Corollary 3.4.6. *Suppose* $\mathrm{conv}(\mathcal{A})$ *is simplicial and includes the origin among its vertices. If for each nonextremal* α *we have that* (i) $\langle c, v \rangle$ *is decreasing in* v_α *and that* (ii) *each* $\langle g_j, v \rangle$ *is increasing in* v_α, *then* $(f, g)_{\mathsf{SAGE}} = (f, g)^\star$.

Proof sketch. The claim that $(f, g)_{\mathsf{SAGE}} = (f, g)^\star$ can be established by a change-of-variables and change-of-constraints argument of the same kind used in the proof of Theorem 3.4.1.

Suffice it to say that rather than using Corollary 3.3.2 to justify removing constraints from the dual without loss of generality, one can simply throw out those constraints to obtain some $(f, g)'$ with $(f, g)' \leq (f, g)_{\mathsf{SAGE}}$. One then shows $(f, g)' = (f, g)^\star$ to sandwich $(f, g)^\star \leq (f, g)' \leq (f, g)_{\mathsf{SAGE}} \leq (f, g)^\star$. $\square$

3.4.5 Finite error in SAGE relaxations

This section's final theorem directly considers SAGE as a relaxation scheme for signomial minimization. It leverages the primal formulation for $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ to establish sufficient conditions under which SAGE relaxations can only exhibit finite error.

Theorem 3.4.7. *Suppose* $\mathrm{conv}(\mathcal{A})$ *contains the origin and there exists an* $\epsilon > 0$ *so that* $(1 + \epsilon)\alpha$ *belongs to* $\mathrm{conv}(\mathcal{A})$ *for all nonextremal* α. *Then* $f = \mathrm{Sig}(\mathcal{A}, c)$ *is bounded below if and only if* $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ *is finite.*

The requirements Theorem 3.4.7 imposes on the Newton polytope are significantly weaker than those found elsewhere in this work. Theorem 3.4.7 is especially notable as we do not know of analogous theorems in the literature on SOS relaxations for polynomial optimization.

Theorem 3.4.7. Let $f = \mathrm{Sig}(\mathcal{A}, c)$ have $f_{\mathbb{R}^n}^\star > -\infty$. We may assume without loss of generality that $0 \in \mathcal{A}$ and that $c_0 = 0$. Use $\mathcal{E} = \{\alpha \in \mathcal{A} : \alpha \text{ nonzero, extremal}\}$ to denote indices of extremal exponents of f, excluding the possibly-extremal zero

---

[1]See Section 3.4 of [13].

vector. The desired claim holds if there exists a positive constant γ so that the translate $\hat{f} = f + \gamma$ is SAGE.

Define $\hat{c} = c + \gamma \delta_0$ as the coefficient vector of $\hat{f}$ in $\mathbb{R}^{\mathcal{A}}$. Because $f^{\star}_{\mathbb{R}^n} > -\infty$ we have $c_\alpha = \hat{c}_\alpha \geq 0$ for every α in $\mathcal{E}$ (Proposition 3.2.1). Let $\mathcal{N} \subset \mathcal{A}$ denote the set of exponents β for which $\hat{c}_\beta < 0$. For each $\beta \in \mathcal{N}$ we define the vector $\hat{c}^{(\beta)}$ in $\mathbb{R}^{\mathcal{A}}$ by

$$\hat{c}^{(\beta)}_\alpha = \begin{cases} \hat{c}_\beta & \text{if } \alpha = \beta \\ \hat{c}_\alpha / |\mathcal{N}| & \text{if } \alpha \in \mathcal{A} \setminus \mathcal{N} \\ 0 & \text{if } \alpha \in \mathcal{N} \setminus \{\beta\} \end{cases} .$$

It is easy to verify that these vectors sum to $\hat{c}$, that each sub-vector $\hat{c}^{(\beta)}_{\setminus \beta}$ is nonnegative, and furthermore that each $\hat{c}^{(\beta)}_0 = \gamma / |\mathcal{N}|$.

We turn to building the corresponding vectors $v^{(\beta)} \in \mathbb{R}^{\mathcal{A}}$. Because $\mathcal{N}$ is contained in $\mathcal{A} \setminus \mathcal{E}$, we have that each $\beta \in \mathcal{N}$ satisfies $(1 + \epsilon)\beta \in \text{conv}(\mathcal{A})$ for some positive ϵ. Therefore each vector β in $\mathcal{N}$ is expressible as a convex combination of extremal exponents and the zero vector. Let $\lambda^{(\beta)} \in \Delta_{\mathcal{E} \cup \{0\}}$ be such a vector of convex combination coefficients (i.e., let $\lambda^{(\beta)}$ satisfy $\beta = [\mathcal{E} \cup \{0\}]^\dagger \lambda^{(\beta)}$ and $\lambda^{(\beta)}_0 > 0$). Given this vector, we set

$$v^{(\beta)}_\alpha = \begin{cases} -1 & \text{if } \alpha = \beta \\ \lambda^{(\beta)}_\alpha & \text{if } \alpha \in \mathcal{E} \cup \{0\} \\ 0 & \text{otherwise} \end{cases} .$$

Each $v^{(\beta)}$ satisfies the inequalities $v^{(\beta)}_{\setminus \beta} \geq 0$, the equations $\mathcal{A}^\dagger v^{(\beta)} = 0$ and $\langle 1, v \rangle = 0$, and has $0 < v^{(\beta)}_0$. The properties $0 < v^{(\beta)}_0$ and $\hat{c}^{(\beta)}_0 = \gamma / |\mathcal{N}|$ ensure that

$$v^{(\beta)}_0 \log \left(v^{(\beta)}_0 / \hat{c}^{(\beta)}_0 \right) \to -\infty \quad \text{as} \quad \gamma \to \infty$$

It follows that there exists a sufficiently large M so that $\gamma \geq M$ implies

$$D(v^{(\beta)}_{\setminus \beta}, e\hat{c}^{(\beta)}_{\setminus \beta}) \leq \hat{c}_\beta \quad \text{for all } \beta \text{ in } \mathcal{N}.$$

Hence for sufficiently large γ, we have $\hat{c}^{(\beta)}$ in $C_{\text{AGE}}(\mathcal{A}, \beta)$ for all β in N – and the result follows. $\square$

3.5   Certifying polynomial nonnegativity

Throughout this section we write $p = \text{Pol}(\mathcal{A}, c)$ to mean that p takes values $p(x) = \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$. We refer to polynomials in this way to reflect our interest in *sparse*

polynomials. Vectors α are sometimes called *terms*, where a term is *even* if α belongs to $(2\mathbb{N})^n$. To a set $\mathcal{A} \subset \mathbb{R}^n$, we associate the sparse nonnegativity cone

$$C_{\mathsf{NNP}}(\mathcal{A}) := \{c \in \mathbb{R}^{\mathcal{A}} : \mathrm{Pol}(\mathcal{A}, c)(x) \geq 0 \text{ for all } x \text{ in } \mathbb{R}^n\}.$$

Beginning with Subsection 3.5.1 we introduce *polynomial SAGE certificates*. We shall see that polynomial SAGE certificates offer a tractable avenue for optimizing over a subset of $C_{\mathsf{NNP}}(\mathcal{A})$, where the complexity depends on $\mathcal{A}$ exclusively through the dimensions n and the cardinality $|\mathcal{A}|$.

Subsection 3.5.2 demonstrates how our study of SAGE signomials yields several corollaries in this new polynomial setting. Perhaps most prominently, Subsection 3.5.2 implies that a polynomial admits a SAGE certificate if and only if it admits a SONC certificate. The qualitative relationship between SAGE and SONC as proof systems is explained in Subsection 3.5.3, and Subsection 3.5.4 addresses how some of our corollaries compare to earlier results in the SONC literature.

In Subsection 3.5.5 we compare polynomial SAGE certificates to the widely-studied SOS certificates. We conclude with Subsection 3.5.6, which outlines how to use SAGE polynomials to obtain a hierarchy for constrained polynomial optimization.

3.5.1 Signomial representatives and polynomial SAGE certificates

To a polynomial $p = \mathrm{Pol}(\mathcal{A}, c)$ we associate the *signomial representative* $q = \mathrm{Sig}(\mathcal{A}, \hat{c})$ with

$$\hat{c}_\alpha = \begin{cases} c_\alpha & \text{if } \alpha \text{ is even} \\ -|c_\alpha| & \text{otherwise} \end{cases}. \tag{3.16}$$

By a termwise argument, we have that if the signomial q is nonnegative on $\mathbb{R}^n$, then the polynomial p must also be nonnegative on $\mathbb{R}^n$. Moving from a polynomial to its signomial representative often entails some loss of generality. For example, the univariate polynomial $p(x) = 1 + x - x^3 + x^4$ never has both "$+x < 0$" and "$-x^3 < 0$," and yet the inner terms appearing in the signomial representative $q(y) = 1 - \exp(y) - \exp(3y) + \exp(4y)$ are both negative.

There is a natural condition $\mathcal{A}$ and the sign pattern of c where passing to the signomial representative is at no loss of generality. Specifically, if there exists a point $x_o \in (\mathbb{R} \setminus \{0\})^n$ where $c_\alpha x_o^\alpha \leq 0$ for all $\alpha \notin (2\mathbb{N})^n$, then $\mathrm{Pol}(\mathcal{A}, c)$ is nonnegative if and only if its signomial representative is nonnegative. We call such polynomials *orthant-dominated*. Checking if a polynomial is orthant-dominated is a simple task. Given $\mathcal{A}$ and c, define $b \in \{0, 1\}^{\mathcal{A}}$ by $b_\alpha = 0$ if $c_\alpha \leq 0$ or α is even,

and $b_\alpha = 1$ if otherwise. Then assuming every $c_\alpha \neq 0$, the polynomial $\mathrm{Pol}(\mathcal{A}, c)$ is orthant-dominated if and only if the system $\mathcal{A}s = b \pmod 2$ has a solution over $s \in \mathbb{F}_2^n$.

We call $p = \mathrm{Pol}(\mathcal{A}, c)$ a *SAGE polynomial* if its signomial representative $q = \mathrm{Sig}(\mathcal{A}, \hat{c})$ is a SAGE signomial. Subsequently, we define a *polynomial SAGE certificate* for $p = \mathrm{Pol}(\mathcal{A}, c)$ as a set of signomial AGE certificates $\{(\hat{c}^{(\beta)}, v^{(\beta)})\}_{\beta \in \mathcal{A}}$ where $\hat{c} := \sum_{\beta \in \mathcal{A}} \hat{c}^{(\beta)}$ defines the signomial representative for p. Because the signomial SAGE cone contains the nonnegative orthant, the cone of coefficients for SAGE polynomials admits the representation

$$C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}) = \{c \in \mathbb{R}^{\mathcal{A}} \; : \; \text{there exists } \hat{c} \text{ in } C_{\mathsf{SAGE}}(\mathcal{A}) \text{ where } \hat{c} \leq c$$
$$\text{and } \hat{c}_\alpha \leq -c_\alpha \text{ for all } \alpha \text{ not in } (2\mathbb{N})^n\}. \qquad (3.17)$$

We use this representation to obtain the following theorem.

Theorem 3.5.1. *Let $L : \mathbb{R}^\ell \to \mathbb{R}^{\mathcal{A}}$ be an injective affine map, identify $\mathcal{A} \subset \mathbb{N}^n$ as an $m \times n$ matrix ($n \leq m$), and let h be a vector in $\mathbb{R}^\ell$. An ϵ-approximate solution to*

$$\inf_{z \in \mathbb{R}^\ell} \{\langle h, z \rangle \; : \; L(z) \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A})\} \qquad (3.18)$$

can be computed in time $O(p(m) \log(1/\epsilon))$ for a polynomial p.

Proof. Throughout the proof we identify $\mathbb{R}^{\mathcal{A}} = \mathbb{R}^m$. We appeal to standard results on interior point methods (IPMs) for conic programming. The task is to show that $C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A})$ can be expressed as a projection of a convex cone "K," which possesses a tractable self-concordant barrier with a complexity parameter ϑ bounded by a polynomial in m. From there, the meaning of "ϵ-approximate" and its relationship to the polynomial "p" depends highly on the details of a given IPM; relevant sources for general conic IPMs include [37, §4] and [105, §5]. In particular we rely on algorithms for optimizing over the *exponential cone* $\mathsf{K}_{\exp} = \mathrm{cl}\{(u, v, w) \; : \; v \exp(u/v) \leq w, v > 0\}$, and defer to [106–108] for formal meanings of "ϵ-approximate" in our context.

For each $i \in [m]$, let $\tilde{\mathcal{A}}_i$ denote the $(m-1) \times n$ matrix with rows given by $\{\alpha_j - \alpha_i \; : \; j \in [m] \setminus i\}$. Next, let M_i denote any matrix whose columns form a basis for the kernel of $\tilde{\mathcal{A}}_i^\top$. We shall index the rows of M_i by $j \in [m] \setminus i$ and say that M_i has "m_i" columns. Finally, define the cone $\mathsf{K}_i = \{(u, v, t) \; : \; u, v \in \mathbb{R}_+^{[m] \setminus i}, D(u, ev) \leq t\}$. In terms of M_i and K_i we can reformulate the α_i^{th} signomial AGE cone as

$$\left\{ c^{(i)} \in \mathbb{R}^m \; : \; \text{some } w^{(i)} \in \mathbb{R}^{m_i} \text{ satisfies } \left(M_i w^{(i)}, \hat{c}_{\setminus i}^{(i)}, \hat{c}_i^{(i)} \right) \in \mathsf{K}_i \right\}.$$

Since K_i can be represented with $m-1$ copies of $\mathsf{K}_{\exp}$ and one linear inequality over $m-1$ additional scalar variables, the preceding display tells us that $C_{\mathsf{SAGE}}(\mathcal{A})$ can be represented with $m(m-1)$ copies of $\mathsf{K}_{\exp}$, m linear inequalities, and $O(m^2)$ scalar auxiliary variables. Combine this with the representation (3.17) to find that the feasible set for (3.18) can be described with $O(m^2)$ exponential cone constraints, $O(m)$ linear inequalities, and $O(\ell + m^2) \in O(m^2)$ scalar variables. As the exponential cone has a tractable self-concordant barrier with complexity parameter $\vartheta_{\exp} = 3$, $C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A})$ has a tractable self-concordant barrier with complexity parameter $O(m^2)$. □

3.5.2 Simple consequences of our signomial results

Subsection 3.5.1 suggested that the signomial SAGE cone is more fundamental than the polynomial SAGE cone. This section serves to emphasize that idea, by showing how our study of the signomial SAGE cone quickly produces results in the polynomial setting. The following corollaries are obtained by viewing Theorems 3.4.1 and 3.4.3 through the lens of orthant-dominance.

Corollary 3.5.2. *If $\mathcal{A}$ induces a simplicial polytope* $\mathrm{conv}(\mathcal{A})$*, and nonextremal exponents are linearly independent mod 2, then* $C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}) = C_{\mathsf{NNP}}(\mathcal{A})$*.*

Corollary 3.5.3. *Suppose $\mathcal{A}$ belonging to $p = \mathrm{Pol}(\mathcal{A}, c)$ can be partitioned into faces where (1) each simplicial face induces an orthant-dominated polynomial with at most two nonextremal terms, and (2) all other faces have at most one nonextremal term. Then p is nonnegative if and only if it is SAGE.*

Unfortunately it is not possible to reduce the dependence of Corollary 3.5.3 on the coefficient vector c of the polynomial p. The obstruction is that taking a signomial representative is not without loss of generality, as the case $\mathcal{A} = [0, 1, 3, 4]$ shows.

To more deeply understand the polynomial SAGE cone it is necessary to study its extreme rays, as wells as its sparsity preservation properties. We now show how this can be done by leveraging Theorems 3.3.1 and 3.3.3 from Section 3.3.

Theorem 3.5.4. *Defining the cone of "AGE polynomials" for exponents $\mathcal{A}$ and index k as*

$$C_{\mathsf{AGE}}^{\mathsf{POLY}}(\mathcal{A}, \boldsymbol{\beta}) := \{c : \mathrm{Pol}(\mathcal{A}, c) \text{ is globally nonnegative, and}$$

$$c_{\backslash \beta} \geq 0, \ c_{\alpha} = 0 \text{ for all } \alpha \neq \beta \text{ where } \alpha \notin (2\mathbb{N})^n\}, \quad (3.19)$$

we have $\sum_{\beta \in \mathcal{A}} C_{\mathsf{AGE}}^{\mathsf{POLY}}(\mathcal{A}, \boldsymbol{\beta}) = C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A})$.

Proof. The inclusion $C_{\text{AGE}}^{\text{POLY}}(\mathcal{A}, \boldsymbol{\beta}) \subset C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A})$ is obvious, since polynomials satisfying (3.19) have AGE signomial representatives. We must show the reverse inclusion $C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A}) \subset \sum_{\boldsymbol{\beta} \in \mathcal{A}} C_{\text{AGE}}^{\text{POLY}}(\mathcal{A}, \boldsymbol{\beta})$.

Given a polynomial $p = \text{Pol}(\mathcal{A}, \boldsymbol{c})$, testing if $\boldsymbol{c}$ belongs to $C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A})$ will reduce to testing if $\hat{\boldsymbol{c}}$ (given by Equation (3.16)) belongs to $C_{\text{SAGE}}(\mathcal{A})$. Henceforth let $\hat{\boldsymbol{c}} \in C_{\text{SAGE}}(\mathcal{A})$ be fixed and set $\mathcal{N} = \{\boldsymbol{\alpha} : \hat{c}_{\boldsymbol{\alpha}} < 0\}$. By Theorem 3.3.1, there exist vectors $\{\hat{\boldsymbol{c}}^{(\boldsymbol{\beta})} \in C_{\text{AGE}}(\mathcal{A}, \boldsymbol{\beta})\}_{\boldsymbol{\beta} \in \mathcal{N}}$ where $\hat{c}_{\boldsymbol{\beta}}^{(\boldsymbol{\beta})} = \hat{c}_{\boldsymbol{\beta}} < 0$ for each $\boldsymbol{\beta}$ and $\hat{c}_{\boldsymbol{\alpha}}^{(\boldsymbol{\beta})} = 0$ for all $\boldsymbol{\alpha} \in \mathcal{N} \setminus \{\boldsymbol{\beta}\}$. The sign patterns here are important: $\hat{\boldsymbol{c}}^{(\boldsymbol{\beta})}$ is supported on $\{\boldsymbol{\beta}\} \cup (\mathcal{A} \setminus \mathcal{N})$, and $\hat{c}_{\boldsymbol{\alpha}}^{(\boldsymbol{\beta})} \geq 0$ for all $\boldsymbol{\alpha}$ in $\mathcal{A} \setminus \mathcal{N}$. By construction of $\hat{\boldsymbol{c}}$, any $\boldsymbol{\alpha}$ in $\mathcal{A} \setminus \mathcal{N}$ satisfies $\boldsymbol{\alpha} \in (2\mathbb{N})^n$. Therefore the carefully chosen vectors $\{\hat{\boldsymbol{c}}^{(\boldsymbol{\beta})}\}_{\boldsymbol{\beta} \in \mathcal{N}}$ define not only AGE signomials, but also *AGE polynomials* $\hat{p}_{\boldsymbol{\beta}} = \text{Pol}(\mathcal{A}, \hat{\boldsymbol{c}}^{(\boldsymbol{\beta})})$. Lastly, for each index $\boldsymbol{\beta} \in \mathcal{N}$ set $\boldsymbol{c}^{(\boldsymbol{\beta})}$ by $c_{\setminus \boldsymbol{\beta}}^{(\boldsymbol{\beta})} = \hat{c}_{\setminus \boldsymbol{\beta}}^{(\boldsymbol{\beta})}$, and $c_{\boldsymbol{\beta}}^{(\boldsymbol{\beta})} = -1 \cdot \text{sign}(c_{\boldsymbol{\beta}}) \cdot \hat{c}_{\boldsymbol{\beta}}^{(\boldsymbol{\beta})}$. The resulting polynomials $p_{\boldsymbol{\beta}} = \text{Pol}(\mathcal{A}, \boldsymbol{c}^{(\boldsymbol{\beta})})$ inherit the AGE property from $\hat{p}_i$, and sum to p. As we have decomposed our SAGE polynomial into an appropriate sum of "AGE polynomials," the proof is complete. $\qquad \square$

Corollary 3.5.5. *Any SAGE polynomial can be decomposed into a sum of AGE polynomials in a manner that is cancellation-free.*

Proof. The cancelation-free decomposition is given constructively in the proof of Theorem 3.5.4. $\qquad \square$

Corollary 3.5.6. *If $\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}}$ generates an extreme ray of $C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A})$, then $\text{supp}\,\boldsymbol{c}$ is either a singleton or a simplicial circuit.*

Proof. In view of Theorem 3.5.4, it suffices to show that for fixed $\boldsymbol{\beta}$ the extreme rays of $C_{\text{AGE}}^{\text{POLY}}(\mathcal{A}, \boldsymbol{\beta})$ are supported on single coordinates, or simplicial circuits. This follows from Theorem 3.3.3, since vectors in $C_{\text{AGE}}^{\text{POLY}}(\mathcal{A}, \boldsymbol{\beta})$ are – up to a sign change on their $\boldsymbol{\beta}^{\text{th}}$ component – in 1-to-1 correspondence with vectors in $C_{\text{AGE}}(\hat{\mathcal{A}}, \boldsymbol{\beta})$, where $\hat{\mathcal{A}}$ is obtained by dropping suitable rows from $\mathcal{A}$. $\qquad \square$

3.5.3 AM/GM proofs of nonnegativity, circuits, and SAGE

In 1989, Reznick defined an *agiform* as any positive multiple of a homogeneous polynomial $f(\boldsymbol{x}) = \sum_{\boldsymbol{\alpha} \in \mathcal{A}} \lambda_{\boldsymbol{\alpha}} \boldsymbol{x}^{\boldsymbol{\alpha}} - \boldsymbol{x}^{\boldsymbol{\beta}}$ where $\boldsymbol{\beta} = \mathcal{A}^{\dagger} \lambda$ for a weighting vector $\lambda \in \Delta_{\mathcal{A}}$. Agiforms have AGE signomial representatives, which follows by plugging $\boldsymbol{v} = \lambda$ into (3.2). Reznick's investigation concerned extremality in the cone of

nonnegative polynomials, and identified a specific subset of simplicial agiforms which met the extremality criterion [14, Theorem 7.1].

Agiform-like functions were later studied by Pantea, Koeppl, and Craciun for analysis of biochemical reaction networks [17, Proposition 3]. Pantea et al. spoke in terms of *posynomials* $f(x) = \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$ where all $c_\alpha \geq 0$; a posynomial f was said to *dominate* the monomial x^β if $x \mapsto f(x) - x^\beta$ was nonnegative on $\mathbb{R}_+^n$. If we adopt the notation where $\Theta(c, \lambda) = \prod_{\alpha \in \mathcal{A}} (c_\alpha / \lambda_\alpha)^{\lambda_\alpha}$, then [17, Theorem 3.6] says that when $\mathcal{A} \cup \{\beta\}$ is a simplicial circuit, monomial domination is equivalent to $1 \leq \Theta(c, \lambda)$ where λ gives the barycentric coordinates for $\beta \in \mathrm{conv}(\mathcal{A})$.
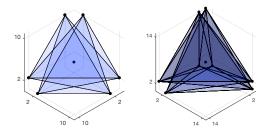
A few years following Pantea et al., Iliman and de Wolff suggested taking sums of nonnegative circuit polynomials, which are globally nonnegative polynomials $f(x) = \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha + b x^\beta$ where $\mathcal{A} \cup \{\beta\}$ form a simplicial circuit [96]. Iliman and de Wolff's Theorem 1.1 states that if all $\alpha \in \mathcal{A}$ are even, f is a circuit polynomial, and $\beta \in \mathrm{conv}(\mathcal{A})$ has barycentric coordinates $\lambda \in \Delta_{\mathcal{A}}$, then f nonnegative if and only if

$$\text{either} \quad |b| \leq \Theta(c, \lambda) \text{ and } \beta \notin (2\mathbb{N})^n \quad \text{or} \quad -b \leq \Theta(c, \lambda) \text{ and } \beta \in (2\mathbb{N})^n. \tag{3.20}$$

It is clear that [96, Theorem 1.1] extends [17, Theorem 3.6], to account for sign changes of $b \cdot x^\beta$ and to impose no scaling on $|b|$. Iliman and de Wolff called $\Theta(c, \lambda)$ the *circuit number* of c, λ.

The approach of taking sums of nonnegative circuit polynomials is now broadly known as "SONC." Prior formulations for the SONC cone work by enumerating every simplicial circuit which could possibly be of use in a SONC decomposition (see [98, §5.2], and subsequently [100, 101]). The circuit enumeration approach is extremely inefficient, as Example 3.5.7 shows an m-term polynomial can contain as many as $2^{(m-1)/2}$ simplicial circuits.

Example 3.5.7. Let d be divisible by 2 and n. Construct a $2n \times n$ matrix $\mathcal{A}$ by setting its rows α_{2i-1} and α_{2i} to distinct points in $\mathbb{N}^n \cap d\Delta_n$ adjacent to $d\delta_i$. Then for large enough d, $\beta = d\mathbf{1}/n$ will be contained in exactly 2^n simplices. Left: $(d, n) = (12, 3)$, and a projection of $(d, n) = (16, 4)$.

Circuit enumeration is not merely a theoretical issue. When using the heuristic circuit-selection technique from [99], Seidler and de Wolff's POEM software package fails to certify nonnegativity of the AGE polynomial $f(x, y) = (x - y)^2 + x^2 y^2$ and moreover only returns a bound $f_{\mathbb{R}^n}^\star \geq -1$ [109].

Of course – Corollary 3.5.6 tells us that a polynomial admits a SAGE certificate if and only if it admits a SONC certificate. This is good news, since Theorem 3.5.1 says we can optimize over this set in time depending polynomially on $|\mathcal{A}|$. In particular, we may avoid SONC's severe problems of circuit enumeration and circuit selection. The qualitative distinction here is that while Pantea et al. and Iliman and de Wolff consider the weights λ as fixed (given by barycentric coordinates), the analogous quantity ν in the SAGE approach is *an optimization variable*. At a technical level, the relative entropy formulation (3.2) affords a joint convexity whereby SAGE can search simultaneously over coefficients $c^{(\beta)}$ and weighting vectors $\nu^{(\beta)}$. As our proof of Theorem 3.5.1 points out, we can be certain that $\mathcal{A}^\dagger \nu^{(\beta)} = \mathbf{0}$ holds in exact arithmetic simply by defining $\nu_{\backslash \beta}^{(\beta)} \leftarrow M_i w^{(i)}$ (for the indicated matrix M_i) and $\nu_\beta^{(\beta)} = -\sum_{\alpha \neq \beta} \nu_\alpha^{(\beta)}$.

3.5.4 Comparison to existing results in the SONC literature

Due to the equivalence of the class of nonnegative polynomials induced by the SAGE and the SONC approaches, some of our results have parallels in the SONC literature.

Corollary 3.5.2 is not stated in the literature, though it may be deduced from [96, Corollary 7.5]. Iliman and de Wolff prove [96, Corollary 7.5] by signomializing $g(x) = f(\exp x)$ and introducing an additional regularity condition so that $\nabla g(x) = \mathbf{0}$ at exactly one $x \in \mathbb{R}^n$. Our proof of Corollary 3.5.2 stems from Theorem 3.4.1, which employs a convex duality argument applicable to constrained signomial optimization problems in the manner of Corollary 3.4.6.

Wang showed that nonnegative polynomials in which at most one term $c_\alpha x^\alpha$ takes on a negative value at some $x \in \mathbb{R}^n$ (either $c_\alpha < 0$ or $\alpha \notin (2\mathbb{N})^n$) are SONC polynomials [100, Theorem 3.9]. This result can be combined with the definition of AGE polynomial given in Theorem 3.5.4 in order to prove a weaker form of Corollary 3.5.6, where all α_i belong to ext $\text{conv}(\mathcal{A})$ or int $\text{conv}(\mathcal{A})$. We emphasize that Corollary 3.5.6 is not responsible for the major efficiency gains of SAGE from Theorem 3.5.1; the SONC formulation in [100, §5] uses $2^{(m-1)/2}$ circuits for the m-term polynomials from Example 3.5.7.

Finally, in a result that was announced contemporaneously to the original submission of the present work, Wang showed that summands in a SONC decomposition of a polynomial $f = \text{Pol}(\mathcal{A}, c)$ may have supports restricted to $\mathcal{A}$ without loss of generality [101, Theorem 4.2]. In light of the equivalence between the class of SAGE polynomials and of SONC polynomials, this result may be viewed as a weaker analog of our Corollary 3.5.5; specifically, [101, Theorem 4.2] shows SONC certificates are sparsity-preserving but it does not provide a cancellation-free decomposition.

The distinctions between our polynomial Corollaries 3.5.2 and 3.5.3 versus our signomial Theorems 3.4.1 and 3.4.3 make clear that polynomial results should not be conflated with signomial results. With the exception of Section 3.5, our setup and results in this work pertain to the class of signomials, which in general can have $\alpha \in (\mathbb{R} \setminus \mathbb{Q})^n$. The developments in the SONC literature only consider polynomials, and employ analysis techniques of an algebraic nature which rely on integrality of exponents in fundamental ways (c.f. [101, Theorem 4.2]). In contrast, our techniques are rooted in convex duality and are applicable to the broader question of certifying signomial nonnegativity.

3.5.5 SAGE and SOS

The SOS approach to polynomial nonnegativity considers polynomials f in n variables of degree $2d$, and attempts to express $f(x) = L(x)^\top P L(x)$ where P is a PSD matrix and $L : \mathbb{R}^n \to \mathbb{R}^{\binom{n+d}{d}}$ is a lifting which maps x to all monomials of degree at-most d evaluated at x [1–3]. The identity $f(x) = L(x)^\top P L(x)$ can be enforced with linear equations on the coefficients of f and the entries of P, so deciding SOS-representability reduces to a semidefinite program.

Because it is extremely challenging to solve semidefinite programs at scale, several modifications to SOS have been proposed to offer reduced complexity. Kojima et. al built on earlier work of Reznick [73] to replace the lifting "L" appearing in the original SOS formulation with a smaller map using fewer monomials [89]. Their techniques had meaningful use-cases, but could fail to perform any reduction in some very simple situations [89, Proposition 5.1]. Subsequently, Waki et. al introduced the *correlative sparsity* heuristic to induce structured sparsity in the matrix variable P [90]. Shortly thereafter Nie and Demmel suggested replacing the standard lifting by a collection of smaller $\{L_i\}_i$, so as to express $f(x) = \sum_i L_i(x)^\top P_i L_i(x)$ with order $\binom{k+d}{d}$ PSD matrices P_i for some $k \ll n$ [91]. Very recently, Ahmadi and Majumdar suggested one use the standard lifting together with a *scaled diagonally*

dominant matrix $\boldsymbol{P}$ of order $\binom{n+d}{d}$;[2] these "SDSOS polynomials" are precisely those polynomials admitting a decomposition as a sum of binomial squares [92].

Each of these SOS-derived works suffers from a drawback that SOS decompositions may require cancellation on coefficients of summands $f_i = g_i^2$ as one recovers $f = \sum_i f_i$. As a concrete example, consider $f(x, y) = 1 - 2x^2 y^2 + x^8/2 + y^8/2$; this polynomial is nonnegative (in fact, AGE) and admits a decomposition as a sum of binomial squares. The trouble is that to decompose f as a sum of binomial squares, the summands $f_i = g_i^2$ require additional terms $+x^4 y^4$ and $-x^4 y^4$. By contrast, SAGE certificates need only involve the original monomials in f, and one may take summand AGE polynomials to be cancellation-free with no loss of generality (Corollary 3.5.5). The SAGE approach also has the benefit of being formulated with a relative entropy program of size $O(m^2)$ (Theorem 3.5.1), while SOS-derived works have complexity scaling exponentially with a polynomial's degree d.

We make two remarks in closing. First, it is easy to verify that every binomial square is an AGE polynomial, and so SAGE can certify nonnegativity of all SDSOS polynomials. Second, it is well known that proof systems leveraging the AM/GM inequality (SAGE among them) can certify nonnegativity of some polynomials which are *not* SOS. A prominent example here is the Motzkin form $f(x, y, z) = x^2 y^4 + x^4 y^2 + z^6 - 3x^2 y^2 z^2$.

3.5.6 Extending SAGE polynomials to a hierarchy

We conclude this section by discussing how to obtain hierarchies for constrained polynomial optimization problems, in a manner which is degree-independent and sparsity preserving. Adopt the standard form (3.15) for minimizing a polynomial f subject to a set of inequality constraints $g(\boldsymbol{x}) \geq 0$ for $g \in G$. Here, all polynomials are over a common set of exponents $\mathcal{A} \in \mathbb{N}^{m \times n}$, with $\boldsymbol{\alpha}_1 = \boldsymbol{0}$ and $m \geq n$. Our development is based on a hierarchy for signomials that is described in [13, §3.3].

Consider operators A and C taking values $A(\mathrm{Pol}(\mathcal{A}, \boldsymbol{c})) = \mathcal{A}$ and $C(\mathrm{Pol}(\mathcal{A}, \boldsymbol{c})) = \boldsymbol{c}$ respectively. We shall say our SAGE polynomial hierarchy is indexed by two parameters: p and q. The parameter p controls the complexity of Lagrange multipliers; when $p = 0$, the Lagrange multipliers are simply $\lambda_i \geq 0$. For general p, the Lagrange multipliers are SAGE polynomials over exponents $\mathcal{A}' := A(\mathrm{Pol}(\mathcal{A}, \boldsymbol{1})^p)$. The parameter q controls the number of constraints in the nonconvex primal prob-

---

[2] A symmetric matrix $\boldsymbol{P}$ is scaled-diagonally-dominant if there exists a diagonal $\boldsymbol{D} > \boldsymbol{0}$ so that $\boldsymbol{DPD}$ is diagonally dominant. Such matrices can be represented as a sum of 2×2 PSD matrices with appropriate zero padding.

lem: $H = \{h_i\}_{i=1}^{k^q}$ are obtained by taking all q-fold products of the g_i. Once the Lagrangian $L = f - \gamma - \sum_{h \in H} h \cdot s_h$ is formed, it will be a polynomial over exponents $\mathcal{A}'' := A(\text{Pol}(\mathcal{A}, \mathbf{1})^{p+q})$. By the minimax inequality we have

$$(f, g)^{(p,q)} := \sup_{\gamma, \{s_h\}_{h \in H}} \{\gamma \ : \ C(L) \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}'') \text{ and}$$

$$C(s_h) \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}') \ \forall \, h \ \in H\} \leq (f, g)^\star.$$

Following Theorem 3.5.1, the above can be solved in time polynomial in m, k for each fixed p, q. As p and q increase, we obtain improved bounds at the expense of an increase in computation. Mirroring [13], one can appeal to representation theorems from the real algebraic geometry literature [110–112] to prove that this hierarchy can provide arbitrarily accurate lower bounds for sparse polynomial optimization problems in which the constraint set is Archimedean (for example, if all variables have explicit finite upper and lower bounds).

Our broader message here – beyond results on convergence to the optimal value of specific hierarchies – is that the above construction qualitatively differs from other hierarchies in the literature, because the optimization problems encountered at every level of our construction *depend only on the nonnegative lattice generated by the original exponent vectors $\mathcal{A}$*. The theoretical underpinnings of this sparsity-preserving hierarchy trace back to the decomposition result given by Theorem 3.3.1. Thus, it is possible to obtain entire families of relative entropy relaxations that are sparsity-preserving, which reinforces our message about the utility of SAGE-based relative entropy optimization for sparse polynomial problems.

3.6 Towards a complete characterization of SAGE versus nonnegativity

We conclude this chapter with a discussion on the extent to which our results tightly characterize the distinction between SAGE and nonnegativity for signomials. This section is split into three parts. In the first part, we describe a process for identifying cases where $C_{\mathsf{SAGE}}(\mathcal{A}) \subsetneq C_{\mathsf{NNS}}(\mathcal{A})$. This process is illustrated with several examples which suggest that our results from Section 3.4 are essentially tight. Subsection 3.6.2 presents a formal conjecture regarding the ways in which our results might be improved, and Subsection 3.6.3 provides a novel dual formulation for when $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$.

3.6.1 Constructing examples of non-equality

In this subsection we treat $\mathcal{A} \in \mathbb{R}^{m \times n}$ as a matrix where the rows are exponent vectors; its transpose is $\mathcal{A}^\top$. For a given $\mathcal{A}$ matrix we are interested in finding a

coefficient vector c so that $f = \mathrm{Sig}(\mathcal{A}, c)$ satisfies $f^{\mathrm{SAGE}}_{\mathbb{R}^n} < f^{\star}_{\mathbb{R}^n}$. If such c exists, then it is evident that $C_{\mathrm{SAGE}}(\mathcal{A}) \neq C_{\mathrm{NNS}}(\mathcal{A})$.

The naive approach to this process would be to construct signomials where the infimum $f^{\star}_{\mathbb{R}^n}$ is known by inspection, to compute $f^{\mathrm{SAGE}}_{\mathbb{R}^n}$, and then to test if the measured value $|f^{\mathrm{SAGE}}_{\mathbb{R}^n} - f^{\star}_{\mathbb{R}^n}|$ is larger than would be possible from rounding errors alone. Unfortunately, it can be quite difficult to construct $\mathcal{A}$ and c where $f^{\star}_{\mathbb{R}^n}$ is apparent and yet $(\mathcal{A}, c)$ are relevant to the conjecture under test.

To address this challenge, we use the "unconstrained SAGE hierarchy" (see [13]) to compute a *sequence* of lower bounds $(f^{(\ell)}_{\mathbb{R}^n})_{\ell \in \mathbb{N}}$. For our purposes, suffice it to say that

$$f^{(\ell)}_{\mathbb{R}^n} := \sup\{\gamma : \mathrm{Sig}(\mathcal{A}, \mathbf{1})^{\ell}(f - \gamma) \text{ is SAGE }\}$$

defines a non-decreasing sequence bounded above by $f^{\star}_{\mathbb{R}^n}$. Note that in particular we have $f^{(0)}_{\mathbb{R}^n} = f^{\mathrm{SAGE}}_{\mathbb{R}^n}$. Thus, while we cannot readily compute $|f^{\mathrm{SAGE}}_{\mathbb{R}^n} - f^{\star}_{\mathbb{R}^n}|$, we *can* compute a few values of $f^{(\ell)}_{\mathbb{R}^n}$ for $\ell > 0$, and check if $|f^{(0)}_{\mathbb{R}^n} - f^{(\ell)}_{\mathbb{R}^n}| \gg 0$.

The remainder of this section probes the sensitivity our earlier theorems' conclusions to their stated assumptions. All computation was performed with a late 2013 MacBook Pro with a 2.4GHz i5 processor, using `CVXPY` [113, 114] as an interface to the conic solver `ECOS` [107, 115].[3] Numerical precision is reported to the farthest decimal point where the primal and dual methods for computing $f^{(\ell)}_{\mathbb{R}^n}$ agree.

Example 3.6.1. We test here whether it is possible to relax the assumption of simplicial Newton polytope in Theorem 3.4.1. Since every Newton polytope in $\mathbb{R}$ is trivially simplicial, the simplest signomials available to us are over $\mathbb{R}^2$. With that in mind, consider

$$\mathcal{A}^{\top} = \begin{bmatrix} 0 & 2 & 1 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 & 1 & 2 \end{bmatrix}.$$

This choice of $\mathcal{A}$ is particularly nice, because were it not for the last element $\alpha_6 = (2, 2)$, we would very clearly have $C_{\mathrm{SAGE}}(\mathcal{A}) = C_{\mathrm{NNS}}(\mathcal{A})$. We tested a few values for c before finding

$$c = (0, 3, -4, 2, -2, 1),$$

which resulted in $f^{(0)}_{\mathbb{R}^n} \approx -1.83333$, and $f^{(1)}_{\mathbb{R}^n} \approx -1.746505595 = f^{\star}_{\mathbb{R}^n}$. Because the absolute deviation $|f^{\mathrm{SAGE}}_{\mathbb{R}^n} - f^{\star}_{\mathbb{R}^n}| \approx 0.08682$ is much larger than the precision to which we solved these relaxations, we conclude that $C_{\mathrm{SAGE}}(\mathcal{A}) \neq C_{\mathrm{NNS}}(\mathcal{A})$ for this choice of $\mathcal{A}$.

---

[3]See `data.caltech.edu/records/1427` for code.

Example 3.6.2. Let us reinforce the conclusion from Example 3.6.1. Applying a 180 degree rotation about the point (1,1) to the rows of $\mathcal{A}$, we obtain

$$\mathcal{A}^{\top} = \begin{bmatrix} 0 & 2 & 0 & 2 & 1 & 2 \\ 0 & 0 & 2 & 2 & 2 & 1 \end{bmatrix}.$$

We then choose the coefficients in a manner informed by the theory developed in Subsection 3.6.3

$$\boldsymbol{c} = (0,\ 1,\ 1,\ 1.9, -2, -2)$$

which subsequently defines $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$. In this case the primal formulation for $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ is infeasible, and so $f_{\mathbb{R}^n}^{(0)} = -\infty$. Meanwhile, the second level of the unconstrained hierarchy produces $f_{\mathbb{R}^n}^{(1)} \approx -0.122211863 = f_{\mathbb{R}^n}^{\star}$. Thus in a very literal sense, the gap $|f_{\mathbb{R}^n}^{\mathsf{SAGE}} - f_{\mathbb{R}^n}^{\star}|$ could not be larger.

We know from Theorem 3.4.3 that any signomial with at most four terms is nonnegative if and only if it is SAGE. It is natural to wonder if in some very restricted setting (e.g. univariate signomials) the SAGE and nonnegativity cones would coincide for signomials with five or more terms; Example 3.6.3 shows this is not true in general.

Example 3.6.3. For $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ with

$$\left[\mathcal{A} \,\middle|\, \boldsymbol{c} \right]^{\top} = \left[\begin{array}{ccccc} 0 & 1 & 2 & 3 & 4 \\ \hline 1 & -4 & 7 & -4 & 1 \end{array} \right].$$

we have $f_{\mathbb{R}^n}^{(0)} \approx -0.3333333$ and $f_{\mathbb{R}^n}^{(1)} \approx 0.2857720944$. Per the affine-invariance properties of the SAGE and nonnegativity cones, this examples shows $C_{\mathsf{SAGE}}(\mathcal{A})$ is a strict subset of $C_{\mathsf{NNS}}(\mathcal{A})$ for every 5×1 matrix $\mathcal{A}$ with equispaced values.

Together, Examples 3.6.1 through 3.6.3 demonstrate there are meaningful senses in which Theorems 3.4.1 through 3.4.3 cannot be improved upon.

3.6.2 A conjecture, under mild regularity conditions

Despite the conclusion in the previous subsection, there *are* settings when we can prove $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$ in spite of $\mathcal{A}$ not satisfying the assumptions of Theorem 3.4.3. For example, one case in which SAGE equals nonnegativity is when

$$\mathcal{A} = \{\boldsymbol{0},\ \boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_n,\ d_1\boldsymbol{\delta}_1, \ldots, d_n\boldsymbol{\delta}_n\}$$

where each d_i belongs to the interval $(0, 1)$. Here one proves equality as follows: for each possible sign pattern of $\boldsymbol{c} \in C_{\mathsf{NNS}}(\mathcal{A})$, there exists a lower dimensional

simplicial face F of $\mathrm{conv}(\mathcal{A})$ upon which we invoke Theorem 3.4.1, and for which the remaining exponents (those outside of F) have positive coefficients. We know that the signomial induced by the exponents outside of F is trivially SAGE, and so by Theorem 3.4.2 we conclude $c \in C_{\mathsf{SAGE}}(\mathcal{A})$. As this holds for all possible sign patterns on c in $C_{\mathsf{NNS}}(\mathcal{A})$, we have $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$. However, this case is rather degenerate, and we wish to exclude it in our discussion via some form of regularity on $\mathcal{A}$.

The most natural regularity condition on $\mathcal{A}$ would be that it admits only the trivial partition. We impose the stronger requirement that every α_i belongs to either $\mathrm{ext}\,\mathrm{conv}(\mathcal{A})$ or $\mathrm{int}\,\mathrm{conv}(\mathcal{A})$. In this setting we have the following corollary of Theorem 3.4.3.

Corollary 3.6.4. *If* $\mathrm{conv}(\mathcal{A})$ *is full dimensional and* $\mathcal{A}$ *has either*

1. *at most one interior exponent, or*

2. $n + 1$ *extreme points and at most two interior exponents*

then $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$.

We also formulated the following conjecture when this work was conducted in 2018. The conjecture has since been resolved in the negative by Forsgård and de Wolff [21, Section 9]. Still, we provide the original statement and supporting examples.

Conjecture 3.6.5. *If* $\mathcal{A}$ *has every* α *in either* $\mathrm{ext}\,\mathrm{conv}(\mathcal{A})$ *or* $\mathrm{int}\,\mathrm{conv}(\mathcal{A})$, *but* $\mathcal{A}$ *does not satisfy the hypothesis of Corollary 3.6.4, then* $C_{\mathsf{SAGE}}(\mathcal{A}) \neq C_{\mathsf{NNS}}(\mathcal{A})$.

Note that when $\mathcal{A}$ satisfies the stated assumptions and and further has $\mathbf{0} \in \mathrm{int}\,\mathrm{conv}\,\mathcal{A}$, Theorem 3.4.7 ensures that $f = \mathrm{Sig}(\mathcal{A}, c)$ can have $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ deviate from $f_{\mathbb{R}^n}^\star$ only by a finite amount. To overcome a potential obstacle posed by this result in the resolution of Conjecture 3.6.5, one can also consider modifying the hypotheses of the conjecture to require that all nonextremal α lie in the relative interior of the Newton polytope. We conclude discussion on this topic with two examples.

Example 3.6.6. Let f be a signomial in two variables with

$$
\left[\mathcal{A} \mid c\right]^\top = \left[\begin{array}{cccccc}
0 & 1 & 0 & 0.30 & 0.21 & 0.16 \\
0 & 0 & 1 & 0.58 & 0.08 & 0.54 \\
\hline
33.94 & 67.29 & 1 & 38.28 & -57.75 & -40.37
\end{array}\right].
$$

Then $f_{\mathbb{R}^n}^{\mathsf{SAGE}} = -24.054866 < f_{\mathbb{R}^n}^{(1)} = -21.31651$. This example provides the minimum number of interior exponents needed to be relevant to Conjecture 3.6.5 in the simplicial case.

Example 3.6.7. Let f be a signomial in two variables with

$$
\left[\mathcal{A} \mid \boldsymbol{c}\right]^{\top} = \left[\begin{array}{cccccc}
0 & 1 & 0 & 2 & 0.52 & 1.30 \\
0 & 0 & 1 & 2 & 0.15 & 1.38 \\
\hline
0.31 & 0.85 & 2.55 & 0.65 & -1.48 & -1.73
\end{array}\right],
$$

then $f_{\mathbb{R}^n}^{\mathsf{SAGE}} = 0.00354263 < f_{\mathbb{R}^n}^{(1)} = 0.13793126$. This signomial has the minimum number of interior exponents needed to be relevant to Conjecture 3.6.5 in the nonsimplicial case.

3.6.3 A dual characterization of SAGE versus nonnegativity

In this section, we provide a general necessary and sufficient dual characterization in terms of certain moment-type mappings for the question of $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$. To establish this dual characterization we use some new notation. Given two vectors $\boldsymbol{u}, \boldsymbol{v}$ the Hadamard product $\boldsymbol{w} = \boldsymbol{u} \circ \boldsymbol{v}$ has entries $w_i = u_i v_i$; this is extended to allow sets in either argument in the same manner as the Minkowski sum. The range of the linear operator $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^{\mathcal{A}}$ is indicated by $\mathcal{A}\mathbb{R}^n$. We extend $\exp : \mathbb{R} \to \mathbb{R}$ first to vectors elementwise and then to sets pointwise.

We begin with the following proposition (proven in the appendix).

Proposition 3.6.8. *If $\boldsymbol{0} \in \mathcal{A}$, then the following are equivalent:*

1. *For every vector $\boldsymbol{c}$, the function $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ satisfies $f_{\mathbb{R}^n}^{\star} = f_{\mathbb{R}^n}^{\mathsf{SAGE}}$.*

2. *$C_{\mathsf{NNS}}(\mathcal{A}) = C_{\mathsf{SAGE}}(\mathcal{A})$.*

3. *$\{\boldsymbol{v} \in \mathbb{R}^{\mathcal{A}} : v_{\boldsymbol{0}} = 1,\ \boldsymbol{v} \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})^{\dagger}\} \subset \mathrm{cl\,conv}\exp(\mathcal{A}\mathbb{R}^n)$.*

Our dual characterization consists of two new sets, both parameterized by $\mathcal{A}$. The first of these sets relates naturally to the third condition in Proposition 3.6.8. Formally, the *moment preimage* of some exponent vectors $\mathcal{A}$ is the set

$$
\mathsf{T}(\mathcal{A}) \coloneqq \log \mathrm{cl\,conv}\exp(\mathcal{A}\mathbb{R}^n).
$$

Here, we extend the logarithm to include $\log 0 = -\infty$ in the natural way. The second set appearing in our dual characterization is defined less explicitly. For a given $\mathcal{A}$,

we say that $\mathsf{S}(\mathcal{A}) \subset \mathbb{R}^{\mathcal{A}}$ is a set of *SAGE-feasible slacks* if $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ has

$$f_{\mathbb{R}^n}^{\mathsf{SAGE}} = \inf\{\langle \boldsymbol{c}, \exp \boldsymbol{y} \rangle : \boldsymbol{y} \text{ in } \mathsf{S}(\mathcal{A}) + \mathcal{A}\mathbb{R}^n\}$$

for every $\boldsymbol{c}$ in $\mathbb{R}^{\mathcal{A}}$.

Theorem 3.6.9. *Let* $\boldsymbol{0} \in \mathcal{A}$ *and* $\mathsf{S}(\mathcal{A})$ *be any set of SAGE-feasible slacks over exponents* $\mathcal{A}$. *Then* $C_{\mathsf{SAGE}}(\mathcal{A}) = C_{\mathsf{NNS}}(\mathcal{A})$ *if and only if* $\mathsf{S}(\mathcal{A}) \subset \mathsf{T}(\mathcal{A})$.

Theorem 3.6.9. To keep notation compact write $\mathsf{U} = \mathcal{A}\mathbb{R}^n$ and $\mathsf{S} = \mathsf{S}(\mathcal{A})$. Also, introduce $\mathsf{W} = \{\boldsymbol{v} : v_{\boldsymbol{0}} = 1, \boldsymbol{v} \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})^{\dagger}\} \subset \mathbb{R}^{\mathcal{A}}$ to describe the feasible set for the dual formulation of $f_{\mathbb{R}^n}^{\mathsf{SAGE}}$. By the supporting-hyperplane characterizations of convex sets, the definitions of S and W imply

$$\mathsf{W} = \mathrm{cl}\,\mathrm{conv}\,\exp(\mathsf{U} + \mathsf{S}).$$

Thus by the equivalence of *1* and *3* in Proposition 3.6.8, it follows that all SAGE relaxations will be exact if and only if $\exp(\mathsf{U} + \mathsf{S}) \subset \mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$. We apply a pointwise logarithm to write the latter condition as $\mathsf{U} + \mathsf{S} \subset \log\mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$.

Now we prove that $\mathsf{T} := \log\mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$ is invariant under translation by vectors in U. It suffices to show that $\exp(\boldsymbol{v} + \mathsf{T}) = \exp\mathsf{T}$ for all vectors $\boldsymbol{v}$ in U. Fixing $\boldsymbol{v}$ in U we have

$$\begin{aligned}
\exp(\boldsymbol{v} + \mathsf{T}) &= \exp(\boldsymbol{v}) \circ \exp(\mathsf{T}) \\
&= \exp(\boldsymbol{v}) \circ \mathrm{cl}\,\mathrm{conv}\,\exp(\mathsf{U}) \\
&= \mathrm{cl}\,\mathrm{conv}\,\exp(\boldsymbol{v} + \mathsf{U}) \\
&= \mathrm{cl}\,\mathrm{conv}\,\exp(\mathsf{U}) = \exp(\mathsf{T})
\end{aligned}$$

as claimed. This translation invariance establishes that $\mathsf{U} + \mathsf{S} \subset \log\mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$ is equivalent to $\mathsf{S} \subset \log\mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$, and in turn that condition *1* of Proposition 3.6.8 holds if and only if $\mathsf{S} \subset \log\mathrm{cl}\,\mathrm{conv}\,\exp\mathsf{U}$. The claim now follows by the equivalence of *1* and *2* in Proposition 3.6.8. $\square$

For constructing sets of SAGE-feasible slacks, one can use a change-of-variables argument similar to that seen in the proof of Theorem 3.4.1. The following proposition shows the outcome of such an argument.

Proposition 3.6.10. *Let* $\Phi : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ *be the extended-real-valued polyhedral mapping with coordinate functions*

$$\Phi_{\beta}(\boldsymbol{s}) = \sup_{\boldsymbol{x} \in \mathbb{R}^n} \left\{ \min_{\alpha \in \mathcal{A}} \{\langle \alpha - \beta, \boldsymbol{x} \rangle + s_{\alpha}\} \right\}.$$

One choice of SAGE-feasible slacks is the polyhedral cone

$$\mathsf{S}(\mathcal{A}) = \left\{ s \in \mathbb{R}^{\mathcal{A}} \mid s_0 = 0,\ \mathbf{0} \le s \le \Phi(s) \right\}.$$

C h a p t e r 4

A NEW APPROACH TO CONSTRAINED SIGNOMIAL NONNEGATIVITY

4.1 Introduction

By now we have seen SAGE and SOS methods for proving function nonnegativity. These methods have transparent applications in bounding the value of an unconstrained minimization problem. Here we consider constrained problems. For K as a proper subset of $\mathbb{R}^n$, we want to take advantage of the following:

$$f_K^\star := \inf_{x \in K} f(x) = \sup\{\gamma \ : \ f - \gamma \text{ is nonnegative on } K\}.$$

So, how can we take global nonnegativity certificates and apply them meaningfully to constrained problems?

One approach takes a page from classical optimization theory. We can replace our constrained problem by an unconstrained problem by forming a suitable Lagrangian L, and we can optimize over Lagrange multipliers subject to a constraint that a shifted Lagrangian $L - \gamma$ is globally nonnegative. This is pragmatic and has been in use in simplified forms (i.e., sans nonnegativity certificates) since the earliest days of optimization. However, traditional Lagrangians often give poor results for nonconvex problems, since it is possible that local minima for L fall far outside K. The *real algebraic approach* to this problem asks more generally for an identity

$$f(x) - \gamma = \sum_{g \in G} g(x)\lambda_g(x) \quad \text{for all} \quad x \in \mathbb{R}^n \tag{4.1}$$

where $g \in G$ are fixed functions known to be nonnegative on K and λ_g are globally nonnegative functions for us to design. Various results from real algebraic geometry say that if $f_K^\star > \gamma$, then under suitable regularity conditions, such an identity exists for sufficiently expressive λ_g.

These algebraic methods are very effective for nonconvex problems. However, the classical optimization literature contains a wealth of methods for strengthening Lagrange dual problems that were developed without real algebraic geometry. One such method is *partial dualization*. The idea is to express an optimization problem's feasible set as

$$K = \{x \ : \ x \in X \text{ and } G(x) \geq 0\}$$

where the constraint $x \in X$ is considered "easy" and $G(x) \geq 0$ is considered "hard." Once we divide our constraints into these categories, we dualize G – that is, we move G into a Lagrangian – and we minimize the Lagrangian over X.

At first blush this only kicks the can down the road. Hasn't the set X just replaced K from before? That depends on what latitude we afford ourselves in choosing X. The real question is

> *How much latitude do we have in choosing a partial dual's "nice set,"*
> *given the tools available to us for handling the final Lagrangian?*

This chapter shows that we have tremendous latitude in applying partial dualization with SAGE signomials, through a concept that we call *conditional SAGE*. We show that when X is a convex set, cones of "X-SAGE signomials" are completely characterized by a relative entropy program involving the support function of X. This result is leveraged to obtain a representation for dual X-SAGE cones, which have a structure enabling a projective solution recovery method for convex relaxations to signomial programs. The qualitative situation is that if X is tractable, then so are primal and dual X-SAGE cones. Our methods therefore provide substantially improved machinery for constrained signomial nonnegativity problems.

This chapter keeps things simple; our goal is to help the reader build familiarity with the basics of conditional SAGE signomials. Chapters 5 and 6 provide deep mathematical investigations into this idea and Chapter 7 applies it to polynomials.

Remark 4.1.1. From this chapter onward, the term "SAGE" has a conceptual meaning. It can refer either to the machinery of conditional SAGE (as introduced here) or the original methods of Chandrasekaran and Shah [13] (which we shall call *ordinary SAGE*). We make limited use of the unqualified term "SAGE" when speaking of specific mathematical objects such as cones.

4.1.1 Related work

The ideas in this chapter are closely related to ordinary SAGE [13] and its immediate relatives: monomial dominating posynomials [17] and nonnegative circuit polynomials [96]. The connections to [17, 96] are mostly genealogical, as the conditional SAGE framework here relies entirely on the convex duality arguments that are unique to [13].

The methods presented here have been used by Wachter, Karaca, Darivianakis, and Charalambous to solve a polynomial optimization problem of interest in power

systems where the decision variables are nonnegative [63]. SOS methods applied to the same problem could not scale to match our methods (see also Chapter 8). Wachter's unpublished Master's thesis covers several topics that are absent from [63], including an exploration of using an arithmetic-geometric inequality over $\mathbb{C}^n$ in an attempt to derive a version of SAGE over the complex numbers [64]. Wachter's particular approach ends with a negative result.

Next, we have Hamza Fawzi's *lifting maps* [116]. In this framework we consider an abstract set S, a vector-valued function $F : \mathsf{S} \to \mathsf{C}$ where $\mathsf{C} \subset \mathbb{R}^m$ is a convex cone, and observe that functions $f(\boldsymbol{x}) = \langle \boldsymbol{c}, F(\boldsymbol{x}) \rangle$ are nonnegative on S provided the coefficient vector $\boldsymbol{c}$ belongs to the dual cone $\mathsf{C}^\dagger$. We do not make a connection to lifting maps in this thesis.

Lastly, we make note of Olga Kuryatnikova's PhD thesis work, which consists broadly of analyzing certificates of nonnegativity using convex analysis rather than real algebraic geometry [117] (see also [118]). Kuryatnikova proves many well-known representation results in real algebraic geometry and creates new ones.

4.1.2   Notation and preliminary definitions

We use the overloaded notation for exponent sets $\mathcal{A} \subset \mathbb{R}^n$ as described in Section 2.1. Recall also that for a vector $\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}}$ and some $\boldsymbol{\beta} \in \mathcal{A}$, we abbreviate $\boldsymbol{c}_{\backslash \boldsymbol{\beta}} := (c_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha} \in \mathcal{A} \backslash \boldsymbol{\beta}}$. Signomials are referred to by $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$. Throughout this chapter, G denotes a vector-valued signomial map used in inequality constraint functions.

For our examples we use `sageopt` and MOSEK [119]. We tun these examples on three machines, two of which are named. Machine W is an HP Z820 workstation, with two 8-core 2.6GHz Intel Xeon E5-2670 processors and 256GB 1600MHz DDR3 RAM. Machine L is a 2013 MacBook Pro, with a dual-core 2.4GHz Intel Core i5 processor and 8GB 1600MHz DDR3 RAM.

4.2   Conditional SAGE certificates of signomial nonnegativity

In this section we show how SAGE certificates for signomial nonnegativity can fully leverage partial dualization, in the sense that any tractable convex set X gives rise to a parameterized and similarly tractable "X-SAGE" nonnegativity cone. The efficient representation of the X-SAGE cones (which we often call "conditional SAGE cones") leads to a practical, principled approach to constrained signomial nonnegativity. For signomial optimization, the most common sets X are of the form $\mathsf{X} = \{\boldsymbol{x} : G(\boldsymbol{x}) \leq \boldsymbol{1}\}$ where G is a posynomial map.

4.2.1   Definition and sparsity preservation

A signomial $\text{Sig}(\mathcal{A}, c)$ is called X-*AGE* if it is nonnegative on X, and at most one component of c is negative. The βth X-AGE cone for signomials over exponents $\mathcal{A} \subset \mathbb{R}^n$ is

$$C_{\mathsf{X}}(\mathcal{A}, \beta) = \left\{ c \in \mathbb{R}^{\mathcal{A}} \ : \ c_{\setminus \beta} \geq 0 \text{ and } \text{Sig}(\mathcal{A}, c)(x) \geq 0 \text{ for all } x \text{ in } \mathsf{X} \right\}.$$

Note that X-AGE cones are defined for arbitrary $\mathsf{X} \subset \mathbb{R}^n$, including nonconvex sets, and convex sets which admit no efficient description. A signomial $\text{Sig}(\mathcal{A}, c)$ is called X-*SAGE* if the coefficient vector c belongs to the Minkowski sum

$$C_{\mathsf{X}}(\mathcal{A}) := \sum_{\beta \in \mathcal{A}} C_{\mathsf{X}}(\mathcal{A}, \beta).$$

We recover the notion of ordinary SAGE by taking $\mathsf{X} = \mathbb{R}^n$.
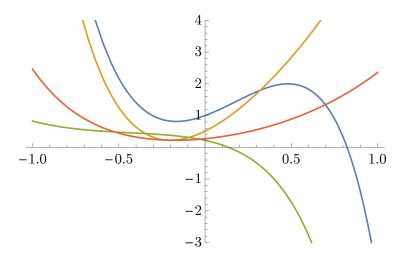
Example 4.2.1. Consider the signomial $f(x) = e^{-3x} + e^{-2x} + 4e^x + e^{2x} - 4e^{-x} - 1 - e^{3x}$ supported on $\mathcal{A} = \{-3, -2, \ldots, 3\}$, together with

$$f_1(x) = 0.88 \cdot e^{-3x} + 0.82 \cdot e^{-2x} + 2.69 \cdot e^x + 0.12 \cdot e^{2x} - 4 \cdot e^{-x},$$
$$f_2(x) = 0.10 \cdot e^{-3x} + 0.15 \cdot e^{-2x} + 0.90 \cdot e^x + 0.12 \cdot e^{2x} - 1, \quad \text{and}$$
$$f_3(x) = 0.02 \cdot e^{-3x} + 0.03 \cdot e^{-2x} + 0.41 \cdot e^x + 0.76 \cdot e^{2x} - e^{3x}.$$

These four signomials are plotted below, with f in blue.

The signomials f_1, f_2, f_3 are evidently nonnegative for $x \leq 0$, each have at most one negative term, are supported on $\mathcal{A}$, and sum to f. We therefore have that f is X-SAGE for $\mathsf{X} = -\mathbb{R}_+$.

It is possible to prove a range of results for X-SAGE signomials without knowing how to check membership in $C_X(\mathcal{A})$. Here is one such result.

Corollary 4.2.2. *Let* $X \subset \mathbb{R}^n$ *be arbitrary. If* $\boldsymbol{c}$ *is a vector in* $C_X(\mathcal{A})$ *with nonempty* $N = \{\boldsymbol{\beta} \in \mathcal{A} : c_\beta < 0\}$, *then there exist vectors* $\{\boldsymbol{c}^{(\beta)}\}_{\beta \in N}$ *satisfying*

$$\boldsymbol{c}^{(\beta)} \in C_X(\mathcal{A}, \boldsymbol{\beta}) \qquad \boldsymbol{c} = \sum_{\beta \in N} \boldsymbol{c}^{(\beta)} \quad \text{and} \quad c_\alpha^{(\beta)} = 0 \text{ for all } \boldsymbol{\alpha} \neq \boldsymbol{\beta} \text{ in } N.$$

One can see Corollary 4.2.2 in action through Example 4.2.1, with $X = -\mathbb{R}_+$ and $\mathcal{A} = \{-3, -2, \ldots, 3\}$. The X-SAGE signomial f had three negative terms and so could be written as a sum of three X-AGE functions. The process of summing f_1, f_2, f_3 to f also resulted in no cancellation of coefficients on the basis functions $\{e^{-x}, 1, e^{3x}\}$. This reduced the dimension of the search space for an X-SAGE decomposition of f from 49 (seven X-AGE cones of dimension seven) to only 12. In large SAGE relaxations, such dimension reduction can reduce computational costs by orders of magnitude.

Corollary 4.2.2 also has a very important theoretical consequence. It tells us there is no loss of generality in restricting X-SAGE decompositions of a signomial $\text{Sig}(\mathcal{A}, \boldsymbol{c})$ to those with X-AGE functions also supported on $\mathcal{A}$. Therefore it is equivalent to say "an X-SAGE function is a sum of X-AGE functions," without making reference to a fixed set of exponent vectors.

Remark 4.2.3. The term "X-AGE" is a *pseudo-acronym*. If an acronym is desired, one might instead call these functions *AGE on* X or *AGE mod* X, and expand "AGE" in the usual way.

4.2.2   Representing conditional SAGE cones

Now we turn to the essential question of how to represent cones of conditional SAGE signomials. The following theorem demonstrates that if X is a tractable convex set, then so is $C_X(\mathcal{A})$.

Theorem 4.2.4. *For exponents* $\mathcal{A} \subset \mathbb{R}^n$, *a vector* $\boldsymbol{\beta} \in \mathcal{A}$, *and a convex set* $X \subset \mathbb{R}^n$ *with support function* $\sigma_X(\lambda) := \sup_{\boldsymbol{x} \in X} \langle \lambda, \boldsymbol{x} \rangle$, *we have*

$$C_X(\mathcal{A}, \boldsymbol{\beta}) = \{\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}} : \text{there exist } \boldsymbol{v} \text{ in } \mathbb{R}^{\mathcal{A}} \text{ and } \boldsymbol{\lambda} \text{ in } \mathbb{R}^n$$
$$\text{satisfying } \mathcal{A}^\dagger \boldsymbol{v} + \boldsymbol{\lambda} = \boldsymbol{0}, \ \langle \boldsymbol{1}, \boldsymbol{v} \rangle = 0,$$
$$\text{and } \sigma_X(\boldsymbol{\lambda}) + D(\boldsymbol{v}_{\backslash \beta}, e\boldsymbol{c}_{\backslash \beta}) \leq c_\beta\}.$$

Theorem 4.2.4 is stated with a support function for maximum generality. From an implementation perspective, it is useful to assume a representation of X. For example, if $X = \{x \ : \ Ax + b \in K\}$ for a matrix A, a vector b, and a convex cone K, then weak duality ensures

$$\sigma_X(\lambda) := \sup\{\langle \lambda, x \rangle \ : \ Ax + b \in K\} \leq \inf\{\langle b, \eta \rangle \ : \ A^\dagger \eta + \lambda = 0, \ \eta \in K^\dagger\}.$$

The above is all we need to construct an inner-approximation of a given AGE cone. For all $X = \{x \ : \ Ax + b \in K\}$, we have

$$\{c \in \mathbb{R}^{\mathcal{A}} : \text{there exist } v \text{ in } \mathbb{R}^{\mathcal{A}} \text{ and } \eta \text{ in } K^\dagger$$
$$\text{satisfying } \mathcal{A}^\dagger v = A^\dagger \eta, \ \langle 1, v \rangle = 0,$$
$$\text{and } D(v_{\backslash \beta}, ec_{\backslash \beta}) + \langle \eta, b \rangle \leq c_\beta\} \subset C_X(\mathcal{A}, \beta).$$

If there exists an x_o where $Ax_o + b$ is in the relative interior of K, then by Slater's condition the reverse inclusion in the preceding expression also holds. Through this approach, we see that if X is relative entropy representable, then so is $C_X(\mathcal{A})$.

This thesis only considers X-SAGE signomials when X is a convex set, however there remains the possibility of using X-SAGE decompositions to certify nonnegativity when X is nonconvex. To give an example of when this is possible, suppose $X = X_1 \cup X_2$ where X_1 and X_2 are convex sets. In this case we trivially have that $C_X(\mathcal{A}, \beta)$ is the intersection of $C_{X_1}(\mathcal{A}, \beta)$ and $C_{X_2}(\mathcal{A}, \beta)$, and so $C_X(\mathcal{A})$ inherits a representation from Theorem 4.2.4.

Remark 4.2.5. The representation of $C := C_{[X_1 \cup X_2]}(\mathcal{A})$ described above is smaller than the standard representation of $C' := C_{X_1}(\mathcal{A}) \cap C_{X_2}(\mathcal{A})$. It is likely the case that $C \subsetneq C'$; it would be interesting to determine the precise nature of this potential gap.

4.2.3 Some observations on the strength of conditional SAGE bounds

To further illustrate that representations of $C_X(\mathcal{A})$ are not required to prove results for X-SAGE signomials, we provide two elementary propositions concerning optimization with conditional SAGE certificates. For $f = \text{Sig}(\mathcal{A}, c)$ with $0 \in \mathcal{A}$, define

$$f_X^{\text{SAGE}} := \sup\{\gamma \ : \ \gamma \text{ in } \mathbb{R}, \ c - \gamma \delta_0 \text{ in } C_X(\mathcal{A})\}$$

so that $f_X^{\text{SAGE}} \leq f_X^\star := \inf\{f(x) : x \text{ in } X\}$.

Proposition 4.2.6. *If $c \geq 0$, then $f = \text{Sig}(\mathcal{A}, c)$ has $f_X^{\text{SAGE}} = f_X^\star$ for all $X \subset \mathbb{R}^n$.*

Proof. The signomial $\tilde{f} = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c} - f_{\mathsf{X}}^{\star}\boldsymbol{\delta_0})$ is nonnegative over X, and its coefficient vector $\boldsymbol{c} - f_{\mathsf{X}}^{\star}\boldsymbol{\delta_0}$ contains at most one negative entry. This implies that $\tilde{f}$ is X-AGE, and hence X-SAGE. $\qquad\square$

Proposition 4.2.7. *If* X *is bounded, then* $f_{\mathsf{X}}^{\mathsf{SAGE}} > -\infty$ *for every signomial* f.

Proof. If X is empty then the result follows by verifying that $C_{\mathsf{X}}(\mathcal{A}) = \mathbb{R}^{\mathcal{A}}$. Consider the case when X is nonempty. In this situation it suffices to prove the result for all f of the form $f(\boldsymbol{x}) = c \exp\langle \boldsymbol{a}, \boldsymbol{x}\rangle$ where $c \neq 0$ and $\boldsymbol{a}$ belongs to $\mathbb{R}^n$. Fixing such $c, \boldsymbol{a}$, the boundedness of X implies the existence of $L \neq 0$ with $\tilde{f}(\boldsymbol{x}) = c \exp\langle \boldsymbol{a}, \boldsymbol{x}\rangle + L$ nonnegative over $\boldsymbol{x}$ in X and $cL < 0$. Since $\tilde{f}$ is nonnegative over X and contains exactly one negative coefficient, we have that $f_{\mathsf{X}}^{\mathsf{SAGE}} \geq -L$. $\qquad\square$

Proposition 4.2.7 shows how convex relaxations based on conditional SAGE certificates respect compactness.

4.2.4 Proofs of main results

Corollary 4.2.2. Let $m = |\mathcal{A}|$ and identify $\mathbb{R}^{\mathcal{A}} = \mathbb{R}^m$. The cones $C_{\mathsf{X}}(\mathcal{A}, \boldsymbol{\beta})$ are of the form $C_i = \{\boldsymbol{c} \in \mathsf{K} : \boldsymbol{c}_{\backslash i} \geq \boldsymbol{0}\}$ where K is the cone of coefficient vectors for X-nonnegative signomials supported on $\mathcal{A}$. Since this cone K contains $\mathbb{R}_+^m$, these C_i satisfy the hypothesis of Lemma 3.3.4. Therefore the proof of the known case with $\mathsf{X} = \mathbb{R}^n$ (Theorem 3.3.1) generalizes immediately. $\qquad\square$

Theorem 4.2.4. For the proof we enumerate $\mathcal{A} = \{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m\}$. Without loss of generality we can take $\boldsymbol{\beta} = \boldsymbol{\alpha}_m$. Let I_{X} denote the indicator function of X, taking values $I_{\mathsf{X}}(\boldsymbol{x}) = 0$ when $\boldsymbol{x} \in \mathsf{X}$ and $I_{\mathsf{X}}(\boldsymbol{x}) = +\infty$ otherwise.

Given $\boldsymbol{c}_{\backslash m} \geq \boldsymbol{0}$, the signomial $f = \mathrm{Sig}(\mathcal{A}, \boldsymbol{c})$ is nonnegative on X if and only if

$$p^{\star} = \inf\{I_{\mathsf{X}}(\boldsymbol{x}) + \textstyle\sum_{i=1}^{m-1} c_i \exp t_i \ : \ \boldsymbol{x} \in \mathbb{R}^n, \ \boldsymbol{t} \in \mathbb{R}^{m-1}, \ \boldsymbol{t} = \boldsymbol{W}\boldsymbol{x}\} \geq -L \qquad (4.2)$$

where $L = c_m$ and $\boldsymbol{W} \in \mathbb{R}^{(m-1)\times n}$ has rows $(\boldsymbol{\alpha}_i - \boldsymbol{\alpha}_m \ : \ i \in [m-1])$. The dual to (4.2) is easily calculated by applying Fenchel duality (c.f. [120]); the result of this process is

$$d^{\star} = \sup\{-\sigma_{\mathsf{X}}(\boldsymbol{\lambda}) - D(\boldsymbol{z}, e\boldsymbol{c}_{\backslash m}) \ : \ \boldsymbol{\lambda} \in \mathbb{R}^n, \boldsymbol{z} \in \mathbb{R}^{m-1}, \ \boldsymbol{W}^{\dagger}\boldsymbol{z} + \boldsymbol{\lambda} = \boldsymbol{0}\}. \qquad (4.3)$$

This problem can be stated in a form closer to the theorem's claim by identifying $\boldsymbol{z} = \boldsymbol{v}_{\backslash \beta}$, and noticing that $\langle \boldsymbol{1}, \boldsymbol{v}\rangle = 0$ implies $v_{\beta} = -\sum_{\alpha \neq \beta} v_{\alpha}$ and $\mathcal{A}^{\dagger}\boldsymbol{v} = \boldsymbol{W}^{\dagger}\boldsymbol{z}$.

When X is nonempty, one may verify that the hypothesis of [120, Corollary 3.3.11] (concerning strong duality) holds for the primal-dual pair (4.2)-(4.3). In particular, $p^\star \geq -L$ holds if and only if $-d^\star \leq L$, and the dual problem attains an optimal solution whenever finite. When X is empty, it is clear that $p^\star = +\infty$, and by taking both λ and ν as zero vectors, we have $d^\star = +\infty$. The result follows. $\square$

4.3 An application in chemical dynamics

Here we consider a problem in chemical dynamics described by Pantea, Koeppl, and Craciun in [17]. The problem concerns a *chemical reaction network* (CRN), encoded as a polynomial map $r : \mathbb{R}_{++}^n \times \mathbb{R}_{++}^m \to \mathbb{R}^n$ that defines a parameterized dynamical system

$$\frac{\mathrm{d}}{\mathrm{d}t} x(t) = r(x(t); k).$$

Here, x is a vector of concentrations of *chemical species*, and k is a vector of *reaction rates* between these species. General background on CRNs can be found in the short tutorial by Yu and Craciun [31]. If each monomial in r involves at most two distinct species x_i, x_j, then we call the reaction network *bi-molecular*.

From r we construct polynomials $p(x; k) = \det \operatorname{Jac} r(x; k)$ where "Jac" computes the Jacobian matrix with respect to x. Pantea et al.'s Theorem 2.1 states that for bi-molecular CRNs, the map $r(\cdot; k)$ is injective on an open convex set $\Omega \subset \mathbb{R}^n$ if and only if $p(x; k) \neq 0$ for all x in Ω. The injectivity property is important because it means the reaction network can have at most one fixed point over $x \in \Omega$.

To see how conditional SAGE can help here, we must reparameterize the polynomial $p(x; k)$ into a signomial $f(\tilde{x}, \tilde{k})$ in $\tilde{x} = \log x$ and $\tilde{k} = \log k$. We are interested in the behavior of $f(\cdot, \tilde{k})$ over $\log \Omega$ for various values of rate constants k. In order for [17, Theorem 2.1] to apply, we need $\log \Omega$ to be a convex set. Simple examples when this happens include when Ω is an open box in $\mathbb{R}_{++}^n$ or the intersection of such a box with a halfspace $\{x : \langle a, x \rangle < 1\}$ defined by some $a \geq 0$.

For our particular example we will work with $\Omega = \mathbb{R}_{++}^n$ and focus on certifying injectivity of $r(\cdot; k)$ for many values of k simultaneously. From a nonnegativity standpoint, we will consider k as *variables* that belong in a box $K \subset \mathbb{R}_{++}^m$. We will apply conditional SAGE to lower bound the minimum of f over $\mathbb{R}^n \times \log K$, and conclude r is injective for all values of $k \in K$ if that lower bound is positive.

Here is the full system we consider:

$$r_1(\boldsymbol{x}; \boldsymbol{k}) = -k_1 x_1 x_2 + k_2 x_4 + 2k_5 x_3 - 2k_6 x_1^2 - k_9 x_1 + k_{10}$$
$$r_2(\boldsymbol{x}; \boldsymbol{k}) = -k_1 x_1 x_2 + k_2 x_4 - k_3 x_2 x_3 + k_4 x_5 - k_{11} x_2 + k_{12}$$
$$r_3(\boldsymbol{x}; \boldsymbol{k}) = -k_3 x_2 x_3 + k_4 x_5 - k_5 x_3 + k_6 x_1^2 - 2k_7 x_3^2 + 2k_8 x_6 - k_{13} x_3 + k_{14}$$
$$r_4(\boldsymbol{x}; \boldsymbol{k}) = k_1 x_1 x_2 - k_2 x_4 - k_{15} x_4 + k_{16}$$
$$r_5(\boldsymbol{x}; \boldsymbol{k}) = k_3 x_2 x_3 - k_4 x_5 - k_{17} x_5 + k_{18}$$
$$r_6(\boldsymbol{x}; \boldsymbol{k}) = k_7 x_3^2 - k_8 x_6 - k_{19} x_6 + k_{20}.$$

(4.4)

We work towards describing $p(\boldsymbol{x}; \boldsymbol{k}) = \det \operatorname{Jac} r(\boldsymbol{x}; \boldsymbol{k})$ in a few steps. First, we note there is a general phenomenon in CRN theory that these polynomials tend to be very sparse and have only a few negative terms (see [17] for discussion and references supporting this claim). For this system, p has 91 positive terms, two negative terms, and the negative terms only depend on $\boldsymbol{x}$ through a bilinear term $x_2 x_3$.

Per the discussion in [17], many of these parameters $\boldsymbol{k}$ have a common physical interpretation, and we can reasonably make the assumption that

$$k_9 = k_{11} = k_{13} = k_{15} = k_{17} = k_{19} = 1.$$

To express the polynomial p, we define the following intermediate expressions in $\boldsymbol{k}$,

$$y_{00} = (1 + k_2)(1 + k_4)(1 + k_5)(1 + k_8),$$
$$y_{10} = (1 + k_8)(k_1 + k_3 + k_1 k_4 + k_1 k_5 + k_2 k_3 + k_1 k_4 k_5),$$
$$y_{01} = (1 + k_2)(k_3 + 4k_7 + k_3 k_5 + k_3 k_8 + 4k_4 k_7 + k_3 k_5 k_8),$$
$$y_{11} = k_1(k_3 + 4k_7 + k_3 k_8 + 4k_4 k_7 - k_3 k_5 - k_3 k_5 k_8),$$
$$y_{20} = k_1 k_3(1 + k_8), \quad y_{02} = 4k_3 k_7(1 + k_2), \quad \text{and} \quad y_{12} = 4k_1 k_3 k_7.$$

Pantea et al. report that

$$p(\boldsymbol{x}; \boldsymbol{k}) = y_{00} + y_{10} x_2 + y_{01} x_3 + y_{20} x_2^2 + y_{11} x_2 x_3 + y_{02} x_3^2 + y_{12} x_2 x_3^2 \quad (4.5)$$
$$+ \text{ terms containing } \boldsymbol{x}\text{-variables other than } x_2 \text{ and } x_3$$

and they analyzed this bivariate polynomial by hand in a few different ways.

In our case we take f as the signomial defined by the expression in $(\boldsymbol{x}, \boldsymbol{y})$ in the top line of (4.5). This signomial has nine variables (two from $\boldsymbol{x}$ and seven from $\boldsymbol{k}$), 49 positive terms, and two negative terms. We consider f over sets $\mathbb{R}^n \times \log \mathsf{K}$ where k_5 is fixed to a certain value and remaining k_i can vary freely over a common interval. We choose these regions because [17] performed a partial by-hand analysis of the case with $k_5 = 10$ and $k_i = 1$ for all $i \neq 5$.
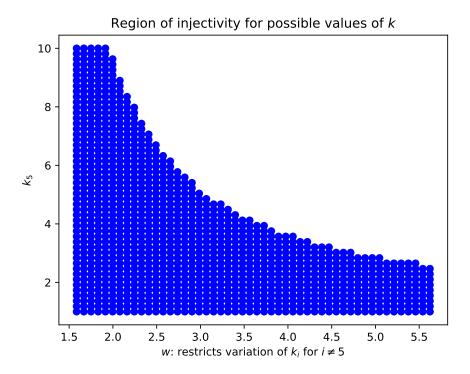
Figure 4.1: Results for testing capacity for multiple equilibria for the chemical reaction network defined by (4.4). The horizontal axis indicates a restriction of k_i for $i \neq 5$ to the interval $[w^{-1}, w]$. The blue dots mark all parameters (w, k_5) in a 50-by-50 grid of $[10^{1/4}, 10^{3/4}] \times [1, 10]$ for which the dynamics map is injective.

Figure 4.1 shows the result of applying conditional SAGE to certify positivity of f over these sets. The smallest positive SAGE bound for these functions was 0.367 and the largest negative SAGE bound was -1.42. In every case of a negative SAGE bound we were able to use the dual SAGE relaxation (described in the next section) to recover a point which proved that f attained a negative value on the relevant domain. Figure 4.1 therefore provides a complete empirical characterization of the parameter choices (w, k_5) for which the CRN defined by (4.4) has capacity for multiple equilibria according to the injectivity test.

We take a moment to address the efficacy of our method. Because $\mathsf{X} := \mathbb{R}^n \times \log \mathsf{K}$ was a rectangular prism, the support function term appearing in the relative entropy representation for $C_{\mathsf{X}}(\mathcal{A})$ could be accommodated by extremely simple linear inequalities. Our `sageopt` python package automated the construction of X given algebraic constraints on $\boldsymbol{k}$ and also performed dimension-reduction based on our Corollary 4.2.2. The effect of `sageopt`'s dimension reduction was to reduce the relative entropy program's number of variables by more than a factor of 20.

We solved these SAGE relaxations using MOSEK 9.2 on a Dell XPS 13 9300, with an Intel Core i7-1065G7 processor (four cores at 1.30GHz) and 16 GB DDR4 RAM (3733 MT/s). For each choice of K, the SAGE relaxation could be solved in less than 0.05 seconds on average with little variation in runtime.

We also approached a handful of these problems (for various K) using SOS-based *Lasserre relaxations* as implemented in GloptiPoly3 [121]. The lowest order Lasserre relaxations we tested took approximately 15 seconds to solve using MOSEK and returned a bound of $-\infty$. A likely cause for vacuous bounds can be found in how the objective function is degree six and the constraint functions (those defining K and restricting $x \geq 0$) are linear. This situation means that the generalized Lagrange multipliers in a degree-six Lasserre relaxation are SOS polynomials of degree at most four, and so the objective function is of higher degree than all other terms in the resulting Lagrangian. Useful bounds can be obtained using degree-six SOS multipliers and hence a degree-eight Lasserre relaxation. For the sets K we tested, the degree-eight Lasserre relaxations took 1200 seconds to solve on average.

4.4 Dual perspectives and solution recovery

Dual SAGE relaxations can be used to recover optimal and near-optimal solutions to signomial programs. For concreteness, we state the simplest such relaxation here. Let $f = \text{Sig}(\mathcal{A}, c)$ be the minimization objective and X be a convex set in $\mathbb{R}^n$. Additionally, consider a signomial map G with coordinate functions $g_i = \text{Sig}(\mathcal{A}, g_i)$ that define a *possibly nonconvex* set $\{x : G(x) \geq 0\}$, and assemble the coefficient vectors g_i into the rows of a matrix G. With this notation,

$$(f, G)_{\mathsf{X}}^{\text{SAGE}} = \inf\{\langle c, v \rangle : v \in C_{\mathsf{X}}(\mathcal{A})^{\dagger}, \, v_1 = 1, \, Gv \geq 0\} \tag{4.6}$$

is a convex relaxation of

$$(f, G)_{\mathsf{X}}^{\star} = \inf\{f(x) : x \in \mathsf{X}, \, G(x) \geq 0\}.$$

By standard rules in convex analysis, the dual SAGE cone is given by the intersection of the constituent dual AGE cones. An expression for the dual AGE cones can be recovered from Theorem 4.2.4 in the case when X is a convex set:

$$C_{\mathsf{X}}(\mathcal{A}, \beta)^{\dagger} = \text{cl}\left\{v \in \mathbb{R}^{\mathcal{A}} : v_{\beta} \log(v/v_{\beta}) \geq [\mathcal{A} - \beta]z_{\beta}\right.$$
$$\left. z_{\beta}/v_{\beta} \in \mathsf{X}, \, v \text{ in } \mathbb{R}_{++}^{\mathcal{A}}, \text{ and } z_{\beta} \text{ in } \mathbb{R}^n\right\}. \tag{4.7}$$

The auxiliary variables "z" appearing in (4.7) are a powerful tool for solution recovery. As long as v is in the *positive orthant* (i.e., if the closure operation in

(4.7) is not relevant for v) then $x := z_\beta / v_\beta$ belongs to X. Beyond taking individual ratios, we note that for any $\mathcal{J} \subset \mathcal{A}$, we have $\left(\sum_{\alpha \in \mathcal{J}} z_\alpha \right) / \left(\sum_{\alpha \in \mathcal{J}} v_\alpha \right) \in$ X. The ability to unconditionally recover X-feasible points by perspective transforms of a dual solution is a powerful feature of the conditional SAGE approach.

In general there remain issues of recovering *optimal* points, and recovering solutions when some constraints cannot be pushed into X. Both of these issues can be resolved if some $x^\star \in$ X satisfies $\exp(\mathcal{A} x^\star) = v$ (as happens when all relative entropy constraints in (4.7) are binding and we meet one additional normalization condition). However it is possible that a SAGE relaxation produces a tight bound, and yet we cannot find a point $x^\star \in$ X with $\exp(\mathcal{A} x^\star) = v$. Therefore it is beneficial to include heuristics in the solution recovery process. Our basic solution recovery algorithm is given below.

Algorithm 1 signomial solution recovery from dual SAGE relaxations.

Input: An objective signomial f and a signomial map G over exponents $\mathcal{A} \subset \mathbb{R}^n$. A vector v in $C_X(\mathcal{A})^\dagger$. Infeasibility tolerance ϵ.

 1: **procedure** SIGSOLUTIONRECOVERY($f, G, \mathcal{A}, v, \epsilon$)
 2:       solutions $\leftarrow$ []
 3:       **for** $\beta \in \mathcal{A}$ **do**
 4:           Recover z_β in $\mathbb{R}^n$ s.t. $v_\beta \log(v / v_\beta) \geq [\mathcal{A} - \beta] z_\beta$ and $z_\beta / v_\beta \in$ X.
 5:           solutions.append(z_β / v_β).
 6:       **if** $\mathcal{A} x \neq \log v$ for all x in solutions **then**
 7:           Compute x_{ls} in argmin$\{ \| \log v - \mathcal{A} x \| : x$ in X$\}$.
 8:           solutions.append(x_{ls}).
 9:       solutions $\leftarrow$ [x in solutions if $G(x) \geq -\epsilon \cdot \mathbf{1}$].
10:       solutions.sort(f, increasing).
11:       **return** solutions.

Assuming (4.7) is used to represent $C_X(\mathcal{A})^\dagger$, Algorithm 1's runtime is dominated by the constrained least-squares problem in Line 7. Note the only projective transformations used in Algorithm 1 are those with index sets $\mathcal{J} = \{\beta\}$; this is due to a present lack of theory for identifying which of the exponentially-many index sets $\mathcal{J} \subset \mathcal{A}$ might be useful for solution recovery. It is highly desirable to develop a systematic theory of solution recovery for dual X-SAGE relaxations, such as that found in Lasserre relaxations for polynomial optimization. In Lasserre relaxations, there are necessary and sufficient conditions for success of solution recovery based on a rank condition for dual variables to SOS multipliers (see [122] for a thorough

treatment of this topic, and [123, Theorem 2.47] for a concise statement of such a result).

4.5 Signomial optimization with convex constraints only

The following problem has appeared in many articles concerning algorithms for signomial programming [77, 80–83].

$$\inf_{\boldsymbol{x} \in \mathbb{R}^3} 0.5 \exp(x_1 - x_2) - \exp x_1 - 5 \exp(-x_2) \tag{4.8}$$

$$\text{s.t. } 100 - \exp(x_2 - x_3) - \exp x_2 - 0.05 \exp(x_1 + x_3) \geq 0$$

$$\exp \boldsymbol{x} - (70, \, 1, \, 0.5) \geq \boldsymbol{0}$$

$$(150, \, 30, \, 21) - \exp \boldsymbol{x} \geq \boldsymbol{0}$$

Problem (4.8) is a good candidate for conditional SAGE relaxations, because each of the seven constraints defines a tractable convex set. The latter six constraints can be represented with six linear inequalities, and the first constraint can be accommodated by three exponential cones and one linear inequality. Separately, problem (4.8) is interesting because Lagrangian approaches perform poorly: regardless of how many products we take of existing constraint functions $G = \{g_1, \ldots, g_7\}$,[1] the "$-5 \exp(-x_2)$" term in the objective will cause Lagrangians $f - \sum_I \lambda_I \prod_{j \in I} g_j$ to be unbounded below for all values of dual variables $\lambda_I \geq 0$.

Now we see how SAGE fares with problem (4.8). Set $\mathsf{X} = \{\boldsymbol{x} : G(\boldsymbol{x}) \geq \boldsymbol{0}\}$; since X is bounded, Proposition 4.2.7 tells us $f_{\mathsf{X}}^{\mathsf{SAGE}}$ is finite. The dual SAGE relaxation can be solved with MOSEK on Machine **L** in 0.01 seconds, and provides us with a lower bound $f_{\mathsf{X}}^{\mathsf{SAGE}} = -147.86 \leq f_{\mathsf{X}}^{\star}$. By running Algorithm 1 on the dual solution, we recover

$$\boldsymbol{x}^{\star} = (5.0106353, \, 3.4011966, -0.4845071) \quad \text{where} \quad f(\boldsymbol{x}^{\star}) = -147.66666.$$

From this solution, we know that the SAGE bound is within 0.13% relative error of the true optimal value. The ability to recover near-optimal solutions even in the presence of a gap $f_{\mathsf{X}}^{\mathsf{SAGE}} < f_{\mathsf{X}}^{\star}$ can be attributed to how conditional SAGE certificates seamlessly integrate with convex duality and partial dualization.

As it happens, the point $\boldsymbol{x}^{\star}$ returned by Algorithm 1 is actually *optimal* for Problem (4.8); to certify this fact, we need stronger SAGE relaxations. We take this opportunity to introduce a *hierarchy* of conditional SAGE relaxations. For a nonnegative

---

[1]See Subsection 8.3.2 for discussion on why taking products of constraints is useful.

integer ℓ, we define

$$f_{\mathsf{X}}^{(\ell)} = \sup\{\gamma \,:\, \mathrm{Sig}(\mathcal{A}, \mathbf{1})^{\ell}(f - \gamma) \text{ is X-SAGE}\}. \tag{4.9}$$

Note the similarity between these bounds and those for unconstrained signomial optimization we used in Subsection 3.6.1. As in that earlier case, is easy to see that these bounds are non-decreasing in ℓ. What is more significant, however, is that these bounds are certain to converge to $f_{\mathsf{X}}^{\star}$ provided X is compact. Table 4.1 shows the result of applying the hierarchy (4.9) to problem (4.8).

level	SAGE bound	**W** time (s)	**L** time (s)
0	-147.85713	0.03	0.01
1	-147.67225	0.05	0.02
2	-147.66680	0.08	0.08
3	-147.66666	0.19	0.26

Table 4.1: SAGE bounds for Example 1, with solver runtime for Machines **W** and **L**. A level-3 bound certifies the level-0 solution as optimal, within relative error 10^{-8}.

Remark 4.5.1. The hierarchy (4.9) was originally proposed in [60], as part of the creation of the "conditional SAGE" concept. The convergence result for the hierarchy in (4.9) was first proven by A. Wang et al. [124] under some regularity conditions, and is extended to a far stronger form in Chapter 6.

4.6 Signomial optimization with convex and nonconvex constraints

This section's example can be found in the 1976 PhD thesis of James Yan [86], where it illustrates signomial programming in the service of structural engineering design. This problem is nonconvex even when written in exponential form; such problems have received limited attention in the engineering design optimization community, largely due to a lack of reliable methods for solving them. We restate the problem here with "generalized polynomials"

$$\inf_{t \in \mathbb{R}_{++}^4} \; 10^4 \, (t_1 + t_2 + t_3) \tag{4.10}$$

$$\text{s.t. } 10^4 + 0.01 \, t_1^{-1} t_3 - 7.0711 \, t_1^{-1} \geq 0$$

$$10^4 + 0.00854 \, t_1^{-1} t_4 - 0.60385 \, (t_1^{-1} + t_2^{-1}) \geq 0$$

$$70.7107 \, t_1^{-1} - t_1^{-1} t_4 - t_3^{-1} t_4 = 0$$

$$10^4 \geq 10^4 \, t_1 \geq 10^{-4} \qquad 10^4 \geq 10^4 \, t_2 \geq 7.0711$$

$$10^4 \geq 10^4 \, t_3 \geq 10^{-4} \qquad 10^4 \geq 10^4 \, t_4 \geq 10^{-4}.$$

Let $X \subset \mathbb{R}^4$ be the feasible set cut out by the eight bound constraints and G be the signomial map that includes all constraints in (4.10) (for a total of 12 inequality constraints). Using the relaxation (4.6) from Section 4.4, we compute $(f, G)_X^{\text{SAGE}} = 14.1423$ in 0.04 seconds of solver time. This bound is very close to the optimal value claimed by Yan [86]. However, Algorithm 1 only returns candidate solutions "$\boldsymbol{x}$" with equality constraint violations $|70.7107 \, t_1^{-1} - t_1^{-1}t_4 - t_3^{-1}t_4| \approx 70$.

To improve our chances of solution recovery, we use the equality constraint to *define* the value $t_4 \leftarrow 70.7107 \, t_3/(t_3 + t_1)$. After clearing the denominator $(t_3 + t_1)$ for inequality constraints involving t_4, we obtain a signomial program in only the variables t_1, t_2, t_3. We compute the analogous value $(f, G)_X^{\text{SAGE}}$ for this configuration and exponentiate the result of Algorithm 1 to recover

$$\boldsymbol{t}_{1:3} = (7.07110 \cdot 10^{-4}, \ 7.07110 \cdot 10^{-4}, \ 10^{-8}), \ t_4 = \frac{70.7107 t_3}{t_1 + t_3}.$$

This solution is feasible up to machine precision, and attains objective matching the 14.142300 SAGE bound. The entire process of solving the SAGE relaxation and recovering the optimal solution takes less than 0.05 seconds on Machine **W**.

C h a p t e r   5

SUBLINEAR CIRCUITS

5.1   Introduction

Given a finite set $\mathcal{A} \subset \mathbb{R}^n$, a signomial supported on $\mathcal{A}$ is a real-linear combination

$$f = \sum_{\alpha \in \mathcal{A}} c_\alpha \mathrm{e}^\alpha \text{ of basis functions } \mathrm{e}^\alpha(\boldsymbol{x}) := \exp\langle \alpha, \boldsymbol{x} \rangle \tag{5.1}$$

with coefficients $\boldsymbol{c} = (c_\alpha)_{\alpha \in \mathcal{A}}$. Signomials are a fundamental class of functions with applications, for example, in chemical reaction networks [15, 16], aircraft design optimization [50, 51], and epidemiological process control [125, 126]; see also [20, 127] and its references for the manifold occurrences in pure and applied mathematics. From a modeling perspective it is often useful to consider signomials under a logarithmic change of variables $\boldsymbol{t} \mapsto f(\log \boldsymbol{t}) = \sum_{\alpha \in \mathcal{A}} c_\alpha \prod_{i=1}^n t_i^{\alpha_i}$ so that for $\mathcal{A} \subset \mathbb{N}^n$, one obtains polynomials over the positive orthant $\mathbb{R}_{++}^n$.

A basic question one might ask of a signomial is when the coefficients $\boldsymbol{c}$ are such that f is globally nonnegative. Framing this question in terms of a signomial's coefficients affords direct connections to polynomials. If the exponent vectors $\mathcal{A}$ are contained in $\mathbb{N}^n$, then f is nonnegative on $\mathbb{R}^n$ if and only if the polynomial $\boldsymbol{t} \mapsto \sum_{\alpha \in \mathcal{A}} c_\alpha \prod_{i=1}^n t_i^{\alpha_i}$ is nonnegative on the nonnegative orthant $\mathbb{R}_+^n$. Deciding such nonnegativity problems is NP-hard in general [36]. However, several researchers have developed sufficient conditions for nonnegativity based on the arithmetic-geometric mean inequality. In contrast to the well-known sums of squares nonnegativity certificates in the polynomial setting (see, e.g., [128, 129]), the techniques based on the arithmetic-geometric inequality are not tied to the notion of a polynomial's degree, and hence also naturally apply to signomials. The earliest results here are due to Reznick [14], with a recent resurgence marked by the works of Pantea, Koeppl, and Craciun [17], Iliman and de Wolff [96], and Chandrasekaran and Shah [13]. Whether considered for signomials or polynomials, such techniques have appealing forms of sparsity preservation in the proofs of nonnegativity [59, 101].

In this chapter, we are concerned with the question of when a signomial supported on exponents $\mathcal{A}$ is nonnegative on a convex set X. We approach this problem through the *conditional SAGE* methodology described in the previous chapter. To review,

the method works as follows: if a signomial f of the form (5.1) has at most one negative coefficient c_β, i.e., if

$$f = \sum_{\alpha \in \mathcal{A} \setminus \beta} c_\alpha \mathrm{e}^\alpha + c_\beta \mathrm{e}^\beta \quad \text{has} \quad c_\alpha \geq 0 \ \text{ for all } \alpha \text{ in } \mathcal{A} \setminus \beta$$

then we may divide out the corresponding basis function e^β to obtain a new signomial $g = \sum_{\alpha \in \mathcal{A}} c_\alpha \mathrm{e}^{[\alpha - \beta]}$ without affecting nonnegativity. Because g is the sum of a signomial with all nonnegative coefficients (a *posynomial*) and a constant, it is convex by construction, and so its X-nonnegativity can be exactly characterized by applying the principle of strong duality in convex optimization. The outcome of this duality argument is that f is X-nonnegative if and only if there exists a dual variable $\nu \in \mathbb{R}^{\mathcal{A}}$ that satisfies a certain relative entropy inequality in ν, c, and the support function of X . Thus, the X-nonnegativity of f can be decided in terms of the subclass of convex optimization called *relative entropy programming*. The X-nonnegative signomials with at most one negative coefficient are called X-*AGE*, and the signomials which decompose into a sum of such functions are called X-*SAGE*. The recognition problem for X-SAGE signomials can likewise be decided by relative entropy programming.

The purpose of this chapter is to undertake a structural analysis of the cones of X-SAGE signomials supported on exponents $\mathcal{A}$. At the outset of this research, our goals were to find counterparts to the many convex-combinatorial properties known for the unconstrained case $\mathsf{X} = \mathbb{R}^n$ [21, 59, 130], and to understand conditional SAGE relative to techniques such as nonnegative circuit polynomials [14, 17, 96]. Towards this end we have introduced an analysis tool of *sublinear circuits* which we call *the* X-*circuits of* $\mathcal{A}$. Our definition of these X-circuits (see Section 5.3) centers on a local, orthant-wise, strict-sublinearity condition for the support function of X composed with $\mathcal{A}$. This construction ensures that the special case of $\mathbb{R}^n$-circuits reduces to the simplicial circuits of the affine-linear matroid induced by $\mathcal{A}$.

We demonstrate that analysis by X-circuits is *extremely effective* in characterizing the structure of X-SAGE cones. One can prove nearly every result in this manuscript assuming nothing of X beyond convexity. Some special treatment is given to the case when X is polyhedral, as this reveals some striking interactions between discrete, convex, and so-called *geometrically convex* or *multiplicatively convex* geometry (see Section 5.5). In a broader sense, a selection of our results have consequences for numerical optimization, such as basis identification in optimization with SAGE

certificates, and a procedure to simplify certain systems of power cone inequalities on the nonnegative orthant.

5.1.1 Main contributions

We begin by introducing some limited notation. The support function of a convex set X, denoted σ_X, is the convex function defined by $\sigma_X(y) = \sup\{\langle y, x \rangle : x \in X\}$. We regard the exponent set $\mathcal{A} \subset \mathbb{R}^n$ as a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^{\mathcal{A}}$. Here we mostly encounter this operator's adjoint $\mathcal{A}^{\dagger}v = \sum_{\alpha \in \mathcal{A}} \alpha v_{\alpha}$. We use $C_X(\mathcal{A})$ to denote the cone of X-SAGE signomials supported on $\mathcal{A}$. For each $\beta \in \mathcal{A}$, we denote the corresponding cone of X-AGE functions by

$$C_X(\mathcal{A}, \beta) = \left\{ f \ : \ f = \sum_{\alpha \in \mathcal{A}} c_{\alpha} e^{\alpha} \text{ is X-nonnegative, } c_{\backslash \beta} \geq 0 \right\} \qquad (5.2)$$

where $c_{\backslash \beta}$ denotes the vector in $\mathbb{R}^{\mathcal{A} \backslash \beta}$ formed by deleting c_{β} from c.

The basic tools for our analysis are *the* X*-circuits of* $\mathcal{A}$ (routinely abbreviated to X-*circuits*). We formulate the X-circuits of $\mathcal{A}$ as nonzero vectors $v^{\star} \in \mathbb{R}^{\mathcal{A}}$ at which the augmented support function $v \mapsto \sigma_X(-\mathcal{A}^{\dagger}v)$ exhibits a strict sublinearity condition (see Definition 5.3.1). We characterize X-circuits as generators of suitable convex cones in $\mathbb{R}^{\mathcal{A}} \times \mathbb{R}$ and usually focus on *normalized* X-circuits $\lambda \in \mathbb{R}^{\mathcal{A}}$, for which the nonnegative entries sum to unity. Theorem 5.3.7 shows that in the polyhedral case, X-circuits are exactly the generators of all one-dimensional elements of a suitable polyhedral fan. A key consequence of Theorem 5.3.7 is that when X is a polyhedron, there are only finitely many normalized X-circuits.

Section 5.4 uses the machinery of X-circuits to understand X-AGE cones. First, we show that if a signomial generates an extreme ray of $C_X(\mathcal{A}, \beta)$, then the dual variable v which certifies its required relative entropy inequality must be an X-circuit (Theorem 5.4.2). Normalized X-circuits λ are then associated to cones of λ-*witnessed AGE functions* $C_X(\mathcal{A}, \lambda)$. The functions in $C_X(\mathcal{A}, \lambda)$ are X-nonnegative signomials admitting a nonnegativity certificate based on a damped power cone inequality in weights λ. Theorem 5.4.4 shows that every X-SAGE function can be written as a sum of λ-witnessed AGE functions for X-circuits λ. In proving this, we formalize the connection between conditional SAGE and prior works for global nonnegativity [14, 17, 96]. Theorem 5.4.4 also motivates a basis identification technique where an approximate relative entropy certificate of $f \in C_X(\mathcal{A})$ may be refined by power cone programming. Combining Theorems 5.3.7 and 5.4.4 yields a corollary that when X is a polyhedron, cones of X-SAGE signomials are (in

principle) power cone representable; this generalizes results by several authors in the unconstrained case [131–134].

Section 5.5 undertakes a thorough analysis of $C_X(\mathcal{A})$. We begin by associating X-circuits λ with affine functions $\phi_\lambda : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}$ given by $\phi_\lambda(\boldsymbol{y}) = \sum_{\alpha \in \mathcal{A}} y_\alpha \lambda_\alpha + \sigma_X(-\mathcal{A}^\dagger \lambda)$. We define the *circuit graph* $G_X(\mathcal{A})$ as the smallest convex cone containing these functions and the constant function $\boldsymbol{y} \mapsto 1$. Upon embedding the affine functions on $\mathbb{R}^{\mathcal{A}}$ into $\mathbb{R}^{\mathcal{A}} \times \mathbb{R}$, Theorem 5.5.4 provides the following identity between the dual SAGE cone $C_X(\mathcal{A})^\dagger$ and the dual circuit graph $G_X(\mathcal{A})^\dagger$

$$C_X(\mathcal{A})^\dagger = \mathrm{cl}\{\exp \boldsymbol{y} \ : \ (\boldsymbol{y}, 1) \in G_X(\mathcal{A})^\dagger\}.$$

Qualitatively, Theorem 5.5.4 says $C_X(\mathcal{A})^\dagger$ is not only convex in the classical sense, but also convex under a logarithmic transformation $S \mapsto \log S = \{\boldsymbol{y} : \exp \boldsymbol{y} \in S\}$. The property of a set being convex under this logarithmic transformation is known by various names, including *log convexity* [135], *geometric convexity* [136, 137], or *multiplicative convexity* [138]. This property has previously been considered in the literature on ordinary SAGE certificates [59, 130], but never in such a systematic way as in our analysis. For example, in view of Theorem 5.5.4 it becomes natural to consider $\Lambda_X^\star(\mathcal{A})$ – *the reduced* X-*circuits of* $\mathcal{A}$ – as the normalized circuits λ for which ϕ_λ generates an extreme ray of the circuit graph. The property of a circuit being "reduced" in this sense is highly restrictive, and yet (by Theorem 5.5.5) we can construct $C_X(\mathcal{A})$ using only λ-witnessed AGE cones as λ runs over $\Lambda_X^\star(\mathcal{A})$. Finally, through a technical lemma (5.5.12), we show how separating hyperplanes in the space of the dual circuit graph may be mapped to separating hyperplanes in the exponentiated space of the dual SAGE cone. This lemma has general applications in simplifying systems of certain power cone constraints on the nonnegative orthant; in our context, it serves as the basis for Theorem 5.5.6, paraphrased below.

If X *is a polyhedron and* $C_X(\mathcal{A})$ *consists of more than just posynomials, then*

$$C_X(\mathcal{A}) = \sum_{\lambda \in \Lambda_X^\star(\mathcal{A})} C_X(\mathcal{A}, \lambda).$$

Moreover, $C_X(\mathcal{A}) \subsetneq \sum_{\lambda \in \Lambda} C_X(\mathcal{A}, \lambda)$ *for every proper subset* $\Lambda \subsetneq \Lambda_X^\star(\mathcal{A})$.

Theorem 5.5.6 provides the most efficient possible description on $C_X(\mathcal{A})$ in terms of power cone inequalities. Its computational implications are addressed briefly in Section 5.7.

Throughout the chapter we illustrate key concepts with the half-line $\mathsf{X} = [0, \infty)$. Specifically, Example 5.3.8 addresses the $[0, \infty)$-circuits of a generic point set $\mathcal{A} \subset \mathbb{R}$, and Example 5.5.7 covers the corresponding *reduced* $[0, \infty)$-circuits. This culminates with a complete characterization of the extreme rays of $C_\mathsf{X}(\mathcal{A})$ for $\mathsf{X} = [0, \infty)$ and $\mathcal{A} \subset \mathbb{R}$ (Proposition 5.6.1).

5.1.2 Related work

Let us begin by introducing some basic concepts from discrete geometry. The *circuits* of the affine-linear matroid induced by $\mathcal{A}$ are the nonzero vectors $v^\star \in \ker(\mathcal{A}^\dagger) \subset \mathbb{R}^\mathcal{A}$ which sum to zero, and whose supports are inclusion minimal among all vectors in $\ker(\mathcal{A}^\dagger)$ that sum to zero. In the SAGE literature one is interested in *simplicial circuits*. These are the circuits $v^\star$ that, upon scaling by a suitable constant, have exactly one negative component. The name *simplicial* is used here because the convex hull of the support $\operatorname{supp} v^\star := \{\alpha \; : \; v^\star_\alpha \neq 0\}$ forms a simplex (possibly of low dimension); exactly one element in $\operatorname{supp} v^\star$ is contained in the relative interior of this simplex. These simplicial circuits are uniquely determined (up to scaling) by their supports. It is therefore common to call a subset $\mathcal{A}' \subset \mathcal{A}$ a simplicial circuit if its convex hull forms a simplex and has a relative interior containing exactly one element of $\mathcal{A}'$.

To situate conditional SAGE in the literature one should look to the close relatives of ordinary SAGE: the *agiforms* of Reznick [14], the *monomial dominating posynomials* of Pantea, Koeppl and Craciun [17], and the *sums of nonnegative circuit* (SONC) polynomials of Iliman and de Wolff [96]. The latter two works determined necessary and sufficient conditions for $\mathbb{R}^n_+$ and $\mathbb{R}^n$-nonnegativity of polynomials supported on a simplicial circuit, based on power cone inequalities in the polynomial's coefficients and circuit vector. In our context, key developments in this area include Wang's discovery of conditions under which a SONC decomposition exists for a given polynomial [100], and Murray, Chandrasekaran, and Wierman's proof that the cone of SONC polynomials can be represented by a projection of a cone of SAGE signomials [59, §5]. From these results it is now understood that although ordinary SAGE and SONC have important differences, the two methods are equivalent to one another for purposes of many structural analyses. Through results in this chapter, we show that the "circuit number" approach of SONC does not generalize to the X-nonnegativity problem in the same manner as SAGE. However, it is possible to describe conditional SAGE in a way which is aesthetically similar to SONC via our λ-witnessed AGE cones.

To appreciate the structural results proven for $C_X(\mathcal{A})$ in this work, it is useful to mention some analogous results proven in the case $X = \mathbb{R}^n$. As a signomial generalization of an earlier result by Reznick [14], Murray, Chandrasekaran, and Wierman have shown that every signomial which generates an extreme ray of $C_{\mathbb{R}^n}(\mathcal{A})$ is supported on either a singleton or a simplicial circuit [59]. Curiously, a given signomial f can be extremal in $C_{\mathbb{R}^n}(\mathcal{A})$ for $\mathcal{A}$ as the support of f, and yet nonextremal in $C_{\mathbb{R}^n}(\mathcal{A}')$ for $\mathcal{A}' \supsetneq \mathcal{A}$. To account for this, Katthän, Naumann, and Theobald introduced the concept of a *reduced circuit*, which they used to obtain a complete characterization of the extreme rays of $C_{\mathbb{R}^n}(\mathcal{A})$ [130]. Subsequently, Forsgård and de Wolff employed regular subdivisions, $\mathcal{A}$-discriminants, and tropical geometry to study how circuits affect the algebraic boundary of the signomial SAGE cone [21]. Our results include direct extensions of the above results by Murray et al. and Katthän et al. to the case of $X \subsetneq \mathbb{R}^n$. For Forsgård and de Wolff's work, our *circuit graph* generalizes their *Reznick cone*.

5.2 Preliminaries

Throughout this chapter, $X \subset \mathbb{R}^n$ is closed, convex, and nonempty, and the set $\mathcal{A} \subset \mathbb{R}^n$ is nonempty and finite. We only consider data $(\mathcal{A}, X)$ where the functions $\{e^\alpha\}_{\alpha \in \mathcal{A}}$ are linearly independent on X. The purpose of this linear independence assumption is to ensure the X-nonnegativity cone is pointed. Equivalently, this assumption ensures the moment cone $\text{co}\{\exp(\mathcal{A}x) \in \mathbb{R}^\mathcal{A} : x \in X\}$ is full-dimensional.

Recall the following result from the previous chapter.

Proposition 5.2.1 (Theorem 1 of [60])**.** *A signomial $\sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ is in $C_X(\mathcal{A}, \beta)$ if and only if there exists a nonzero vector $v \in \mathbb{R}^\mathcal{A}$ that satisfies*

$$\langle \mathbf{1}, v \rangle = 0 \quad and \quad \sigma_X(-\mathcal{A}^\dagger v) + D(v_{\backslash \beta}, ec_{\backslash \beta}) \leq c_\beta. \tag{5.3}$$

The larger goal of this chapter is to reveal the additional structure in the X-SAGE cones $C_X(\mathcal{A})$ that is not immediately apparent from Proposition 5.2.1. From the case $X = \mathbb{R}^n$, the additional structure concerned the supports of signomials that generate extreme rays of $C_{\mathbb{R}^n}(\mathcal{A}, \beta)$ or $C_{\mathbb{R}^n}(\mathcal{A})$. In this context it is standard to use the term *simplicial circuit* in the sense of subsets $\mathcal{A}' \subset \mathcal{A}$. Specifically, $\mathcal{A}' \subset \mathcal{A}$ is a simplicial circuit if it is a minimal affinely dependent set and $\text{conv } \mathcal{A}$ has $|\mathcal{A}'| - 1$ extreme points. This definition of circuits in terms of these subsets $\mathcal{A}' \subset \mathcal{A}$ is equivalent to the definition involving numeric vectors $v^\star \in \mathbb{R}^\mathcal{A}$ (see [21]).

Proposition 5.2.2 (Theorem 5 of [59])**.** *Let $\beta \in \mathcal{A}$. A signomial $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ belongs to $C_{\mathbb{R}^n}(\mathcal{A}, \beta)$ if and only if it can be written as a finite sum $f = \sum_{i=1}^{k} f^{(i)}$ of signomials*

$$f^{(i)} = \sum_{\alpha \in \mathcal{A}} c_\alpha^{(i)} e^\alpha \in C_{\mathbb{R}^n}(\mathcal{A}, \beta), \quad 1 \leq i \leq k,$$

such that the supports $\{\alpha \in \mathcal{A} : c_\alpha^{(i)} \neq 0\}$ are either singletons or simplicial circuits.

Of course, in view of Definition 2.1, Proposition 5.2.2 tells us every $f \in C_{\mathbb{R}^n}(\mathcal{A})$ similarly decomposes into AGE functions supported on singletons and simplicial circuits.

Revealing the full structure of conditional SAGE cones requires consideration to more than just a signomial's support. Therefore, thinking in terms of affine-linear circuits as subsets $A \subset \mathcal{A}$ will not suit our purposes. The following definition codifies our convention of considering affine-linear circuits as numeric vectors.

Definition 5.2.3. *A nonzero vector $v^\star \in \{v \in \mathbb{R}^{\mathcal{A}} : \langle 1, v \rangle = 0\}$ in the kernel of the linear operator $v \mapsto \mathcal{A}^\dagger v = \sum_{\alpha \in \mathcal{A}} \alpha v_\alpha$ is called an $\mathbb{R}^n$-circuit if it is minimally supported and has exactly one negative component.*

It is possible that a given $\mathcal{A}$ has no $\mathbb{R}^n$-circuits, but then every $\alpha \in \mathcal{A}$ would be an extreme point of conv $\mathcal{A}$. This is a degenerate case that results in $C_{\mathbb{R}^n}(\mathcal{A})$ containing only posynomials, but we still give consideration to this possibility throughout the chapter. In the language of Definition 5.2.3, we combine Propositions 5.2.1 and 5.2.2 to obtain the following formulation.

Proposition 5.2.4 (Theorem 4.4 of [21])**.** *Let $\beta \in \mathcal{A}$. A signomial $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ belongs to $C_{\mathbb{R}^n}(\mathcal{A}, \beta)$ if and only if there exist $k \geq 0$ and signomials $f^{(i)} = \sum_{\alpha \in \mathcal{A}} c_\alpha^{(i)} e^\alpha \in C_{\mathbb{R}}(\mathcal{A}, \beta)$, $1 \leq i \leq k$, with $f = \sum_{i=1}^{k} f^{(i)}$ and such that for any signomial $f^{(i)}$ which is not supported on a singleton, there exists an $\mathbb{R}^n$-circuit $v^{(i)} \in \mathbb{R}^{\mathcal{A}}$ with $D(v_{\backslash \beta}^{(i)}, e c_{\backslash \beta}^{(i)}) \leq c_\beta^{(i)}$.*

The relative interior of a convex set S is its interior under the topology induced by its affine hull (the smallest affine space containing S). A face of a convex set $S \subset \mathbb{R}^n$ is any closed convex $F \subset S$ with the following property: if the line segment $[s_1, s_2] := \{\lambda s_1 + (1 - \lambda)s_2 : 0 \leq \lambda \leq 1\}$ is contained in S and the relative interior of $[s_1, s_2]$ hits F, then the entirety of $[s_1, s_2]$ is contained in F. We sometimes write $F \trianglelefteq S$ to indicate that F is a face of S.

A vector $\boldsymbol{v}$ is called an edge generator of a convex cone K if $\{\lambda \boldsymbol{v} \; : \; \lambda \geq 0\}$ is an extreme ray of K. The polar of a convex cone K is $\mathsf{K}^{\circ} := -\mathsf{K}^{\dagger}$. Convex sets $\mathsf{S} \subset \mathbb{R}^{n}$ have convex induced cones $\mathrm{indco}(\mathsf{S}) := \mathrm{cl}\{(\boldsymbol{s}, \mu) \; : \; \mu > 0, \, \boldsymbol{s}/\mu \in \mathsf{S}\} \subset \mathbb{R}^{n+1}$ and recession cones $\mathrm{rec}\,\mathsf{S} := \{\boldsymbol{z} \; : \; \exists \boldsymbol{s} \in \mathsf{S} \text{ such that } \boldsymbol{s} + \lambda \boldsymbol{z} \in \mathsf{S} \; \forall \, t \geq 0\}$.

We call a set a polyhedron if it can be represented by the intersection of finitely many half-spaces; polytopes are the bounded polyhedra.

We sometimes use $\boldsymbol{a}^{\top}\boldsymbol{b}$ to denote the standard inner product of two vectors $\boldsymbol{a}, \boldsymbol{b}$ in a common space.

5.3 Sublinear circuits induced by a point set

We begin this section with a functional analytic definition for the X-circuits of a point set $\mathcal{A}$, generalizing $\mathbb{R}^{n}$-circuits to a constrained setting. After revealing various elementary properties and discussing some examples, we characterize X-circuits in more geometric terms in Theorems 5.3.6 and 5.3.7. In particular the latter theorem interprets X-circuits in terms of normal fans when X is a polyhedron. In Example 5.3.8, we determine the $[0, \infty)$-circuits of a univariate support set $\mathcal{A} \subset \mathbb{R}$; the example is developed further in Section 5.5 and culminates in a theorem completely characterizing the extreme rays of the resulting cone $C_{[0,\infty)}(\mathcal{A})$ in Section 5.6.

The derivations in this section are purely combinatorial and convex-geometric, and make no mention of signomials. However, the definition of X-circuits is ultimately chosen to prepare for studying X-SAGE cones, and in particular it relates to distinguished vectors $\boldsymbol{v} \in \mathbb{R}^{\mathcal{A}}$ that might satisfy (5.3) for certain $\boldsymbol{c} \in \mathbb{R}^{\mathcal{A}}$. Note that (5.3) has an implicit constraint $\boldsymbol{v}_{\backslash \beta} \geq \boldsymbol{0}$ arising from our extended-real-valued definition of relative entropy. To avoid dependence on relative entropy in this section, we frame our discussion of X-circuits in terms of cones

$$\mathsf{N}_{\beta} = \{\boldsymbol{v} \in \mathbb{R}^{\mathcal{A}} \; : \; \boldsymbol{v}_{\backslash \beta} \geq \boldsymbol{0}, \; \langle \boldsymbol{1}, \boldsymbol{v} \rangle = 0\} \tag{5.4}$$

for vectors $\boldsymbol{\beta} \in \mathcal{A}$.

Definition 5.3.1. *A vector $\boldsymbol{v}^{\star} \in \mathsf{N}_{\beta}$ is an X-circuit of $\mathcal{A}$ (or simply, an X-circuit) if (1) it is nonzero, (2) $\sigma_{\mathsf{X}}(-\mathcal{A}^{\dagger}\boldsymbol{v}^{\star}) < \infty$, and (3) it cannot be written as a convex combination of two non-proportional $\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)} \in \mathsf{N}_{\beta}$, for which $\boldsymbol{v} \mapsto \sigma_{\mathsf{X}}(-\mathcal{A}^{\dagger}\boldsymbol{v})$ is linear on $[\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}]$.*

The third condition is equivalent to strict sublinearity of $\boldsymbol{v} \mapsto \sigma_{\mathsf{X}}(-\mathcal{A}^{\dagger}\boldsymbol{v})$ on any line segment in N_{β} that contains $\boldsymbol{v}^{\star}$, except for the trivial line segments which generate

a single ray. The central importance of the sublinearity condition leads us to refer to X-circuits also as *sublinear circuits*; the latter term is helpful in remembering the definition early in our development.

Remark 5.3.2. In the special case $X = \mathbb{R}^n$, condition (2) simplifies to $\mathcal{A}^\dagger v = \mathbf{0}$. In conjunction with the definition of N_β, this shows that the special case $X = \mathbb{R}^n$ of Definition 5.3.1 matches exactly with Definition 5.2.3 of $\mathbb{R}^n$-circuits.

Conceptually, Definition 5.3.1 indicates that X-circuits are essential in capturing the behavior of the augmented support function $v \mapsto \sigma_X(-\mathcal{A}^\dagger v)$ on the given N_β. While developing this concept formally it is convenient for us to enumerate the *positive support* $v^+ := \{\alpha \,:\, v_\alpha > 0\}$, and to identify the unique index $v^- := \beta \in \mathcal{A}$ where $v_\beta < 0$. Note that positive homogeneity of the support function tells us that the property of being a sublinear circuit is invariant under scaling by positive constants. A sublinear circuit is *normalized* if its unique negative term v_β has $v_\beta = -1$, in which case we usually denote it by the symbol λ rather than v. We can normalize a given sublinear circuit by taking the ratio with its infinity norm $\lambda = v/\|v\|_\infty$, because $\|v\|_\infty = |v_\beta|$ for all vectors $v \in N_\beta$.

Example 5.3.3. *(The conic case.)* It is straightforward to determine which $v \in N_\beta$ are X-circuits of $\mathcal{A}$ when X is a cone. In such a setting, the support function of X can only take on the values zero and positive infinity. Hence, $v \mapsto \sigma_X(-\mathcal{A}^\dagger v)$ is trivially linear over all of $V_\beta := \{v \in N_\beta \,:\, \sigma_X(-\mathcal{A}^\dagger v) < \infty\}$. The set V_β is a cone, and reformulating $\sigma_X(-\mathcal{A}^\dagger v) = 0$ as $v \in (\mathcal{A}X)^\dagger$ gives

$$V_\beta = (\ker(\mathcal{A}^\dagger) + \mathcal{A}^{-1}X^\dagger) \cap N_\beta,$$

where $\mathcal{A}^{-1}$ denotes the pseudo-inverse of $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^{\mathcal{A}}$. Therefore, the X-circuits $v \in N_\beta$ are precisely the edge generators of $(\ker(\mathcal{A}^\dagger) + \mathcal{A}^{-1}X^\dagger) \cap N_\beta$.

Regarding again the special case $X = \mathbb{R}^n$ from this conic perspective, we have $X^\dagger = \{\mathbf{0}\}$, so $\mathcal{A}^{-1}X^\dagger = \{\mathbf{0}\}$, and $\ker(\mathcal{A}^\dagger) + \mathcal{A}^{-1}X^\dagger = \ker(\mathcal{A}^\dagger)$, which implies $V_\beta = \ker(\mathcal{A}^\dagger) \cap N_\beta$. It is easily shown that edge generators of $\ker(\mathcal{A}^\dagger) \cap N_\beta$ are precisely those $v \in \ker(\mathcal{A}^\dagger) \cap N_\beta \setminus \{\mathbf{0}\}$ for which $v^+ = \{\alpha \,:\, v_\alpha > 0\}$ are affinely independent, which recovers the matroid-theoretic notion of affine-linear simplicial circuits from the point of view of subsets $\mathcal{A}' \subset \mathcal{A}$.

The following proposition shows that the affine-independence property is a necessary condition for all sublinear circuits. The proposition provides insight because it shows an X-circuit v with $X \subset \mathbb{R}^n$ is restricted to $|\operatorname{supp} v| \le n + 2$.

Proposition 5.3.4. *If $v^\star \in N_\beta$ is an X-circuit, then its positive support $(v^\star)^+ = (\text{supp } v^\star) \setminus \beta$ is affinely independent.*

Proof. From a fixed $v^\star \in N_\beta$, we construct the vector $z = -\mathcal{A}^\dagger v^\star$ and the set

$$U = \{v \in N_\beta \,:\, -\mathcal{A}^\dagger v = z, \, v_\beta = v_\beta^\star\}.$$

The function $v \mapsto \sigma_X(-\mathcal{A}^\dagger v)$ is a constant and equal to $\sigma_X(z)$ on U, and so in order for $v^\star$ to be an X-circuit, it must be a vertex of the polytope U. The set U is in 1-to-1 correspondence with

$$W = \{w \in \mathbb{R}_+^{\mathcal{A} \setminus \beta} \,:\, \textstyle\sum_{\alpha \in \mathcal{A} \setminus \beta}(\beta - \alpha)w_\alpha = z, \, \langle \mathbf{1}, w \rangle = -v_\beta^\star\}$$

by identifying $w = v_{\setminus \beta}$. We can express W in more illuminating matrix notation by defining M as the matrix with columns of the form "$(\beta - \alpha, 1)$" indexed by $\alpha \in \mathcal{A} \setminus \beta$. Specifically, we have $W = \{w \in \mathbb{R}_+^{\mathcal{A} \setminus \beta} \,:\, Mw = (z, -v_\beta^\star)\}$.

Basic polyhedral geometry tells us that all vertices $w^\star$ of W use an affinely independent set of columns from M. Furthermore, a given set of columns from M is affinely independent if and only if the corresponding indices of the columns (as vectors $\alpha \in \mathcal{A} \setminus \beta$) are affinely independent. Since the correspondence between $v \in U$ and $w \in W$ preserves extremality, the vertices of U have affinely independent positive support v^+. $\qquad\square$

The converse of Proposition 5.3.4 is not true. This is to say: not every vector $v \in N_\beta$ with affinely independent v^+ is an X-circuit.

Example 5.3.5. Let $\mathcal{A} \subset \mathbb{R}^2$ consist of the three points $\alpha_1 = (0, 0)$, $\alpha_2 = (1, 0)$, and $\alpha_3 = (0, 1)$, and fix $X = \{x \in \mathbb{R}^2 \,:\, x \geq u\}$ for some point $u \in \mathbb{R}^2$. The vector $v^\star = (-2, 1, 1)$ has $(v^\star)^- = \alpha_1 = (0, 0)$, and $(v^\star)^+ = \{\alpha_2, \alpha_3\} = \{(1, 0), (0, 1)\}$ is affinely independent. Considering $v^{(1)} = (-2, 2, 0)$ and $v^{(2)} = (-2, 0, 2)$, we have $v^\star = \frac{1}{2}(v^{(1)} + v^{(2)}) \in \text{ri } L$ for $L := [v^{(1)}, v^{(2)}]$. Moreover, the mapping $v \mapsto \sigma_X(-\mathcal{A}^\dagger v)$ is linear on L, because for any $\mu_1, \mu_2 \geq 0$ with $\mu_1 + \mu_2 = 1$ we have

$$\sigma_X(\mathcal{A}^\dagger(-\mu_1 v^{(1)} - \mu_2 v^{(2)})) = \sigma_X((-2\mu_1, -2\mu_2)) = -2\mu_1 u_1 - 2\mu_2 u_2$$
$$= \sigma_X((-2\mu_1, 0)) + \sigma_X((0, -2\mu_2)).$$

The last equality is true since $(1, 1) \in \mathbb{R}^2$ maximizes both the objective functions $x \mapsto (-2\mu_1, 0)^\top x$ and $x \mapsto (0, -2\mu_2)^\top x$ on X.[1]

---

[1] Recall, $a^\top b$ is the dot product of vectors a, b.

With the basic exercise of Example 5.3.5 complete, we turn to characterizing sublinear circuits in full generality.

Theorem 5.3.6. *Fix $\beta \in \mathcal{A}$. The convex cone generated by*

$$\mathsf{T} = \{ (v, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v)) \,:\, v \in \mathsf{N}_\beta, \, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v) < \infty\}$$

is pointed and closed. A vector $v^\star \in \mathsf{N}_\beta$ is an X-circuit of $\mathcal{A}$ if and only if $(v^\star, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v^\star))$ is an edge generator for $\mathrm{co}\,\mathsf{T}$.

Proof. Let Q denote the closed convex set $\mathsf{Q} = \{v \,:\, v \in \mathsf{N}_\beta, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v) < \infty\}$. The claim of the theorem is trivially true if $\mathsf{Q} = \{\mathbf{0}\}$, in which case there are no X-circuits $v \in \mathsf{N}_\beta$ and $\mathrm{co}\,\mathsf{T} = \{(\mathbf{0}, 0)\}$ has no extreme rays. We therefore assume for the duration of the proof that Q contains a nonzero vector.

We turn to showing $\mathrm{co}\,\mathsf{T}$ is closed and pointed, particularly beginning with pointedness. For this, observe $\mathrm{co}\,\mathsf{T} \subset \mathsf{N}_\beta \times \mathbb{R}$. Since N_β contains no lines, there are no lines in $\mathrm{co}\,\mathsf{T}$ of the form (v, τ) with $v \neq \mathbf{0}$. Meanwhile, we know that the line spanned by $(\mathbf{0}, 1)$ cannot be contained in $\mathrm{co}\,\mathsf{T}$, since $\sigma_\mathsf{X}(-\mathcal{A}^\dagger \mathbf{0}) = 0$. Now we turn to closedness of $\mathrm{co}\,\mathsf{T}$. Since $\mathsf{Q} \subset \mathsf{N}_\beta$, we can normalize Q against $\{v \,:\, v_\beta = -1\}$ to obtain a compact set

$$\mathsf{Q}_1 := \{\lambda \,:\, \lambda \in \mathsf{Q}, \lambda_\beta = -1\}$$

that satisfies $\mathsf{Q} = \mathrm{co}\,\mathsf{Q}_1$. From Q_1 we construct

$$\mathsf{S}_1 = \{(\lambda, \sigma_\mathsf{X}(-\mathcal{A}^\dagger \lambda)) \,:\, \lambda \in \mathsf{Q}_1\}.$$

The set S_1 inherits compactness from Q_1 (by continuity of $\lambda \mapsto \sigma_\mathsf{X}(-\mathcal{A}^\dagger \lambda)$), and the convex hull $\mathsf{S}_2 = \mathrm{conv}\,\mathsf{S}_1$ inherits compactness from S_1 (as the convex hull of a compact set is compact). It is evident that S_2 does not contain the zero vector, and so by [65, Corollary 9.6.1] we have that $\mathrm{co}\,\mathsf{S}_2$ is closed. We finish this phase of the proof by identifying $\mathrm{co}\,\mathsf{T} = \mathrm{co}\,\mathsf{S}_2$.

At this point we have that $\mathrm{co}\,\mathsf{T}$ is the convex hull of its extreme rays; it remains to determine the nature of these extreme rays. Since T is a generating set for $\mathrm{co}\,\mathsf{T}$ and contains only vectors of the form $(v, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v))$, every edge generator of $\mathrm{co}\,\mathsf{T}$ is given by a nonzero vector $(v^\star, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v^\star))$ for appropriate $v^\star$. It is clear that $v^\star$ must be an X-circuit in order for $(v^\star, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v^\star))$ to be an edge generator of $\mathrm{co}\,\mathsf{T}$. The harder direction is to show that $v^\star$ being an X-circuit is *sufficient* for $(v^\star, \sigma_\mathsf{X}(-\mathcal{A}^\dagger v^\star))$ to be an edge generator for $\mathrm{co}\,\mathsf{T}$.

To handle this direction, begin by defining an affinely independent set $V = \{v^{(i)}\}_{i=1}^{\ell}$ and a vector θ in the relative interior of Δ_{ℓ}, where $v^{\star} = \sum_{i=1}^{\ell} \theta_i v^{(i)}$ and

$$\sigma_X(-\mathcal{A}^{\dagger} v^{\star}) = \sum_{i=1}^{\ell} \theta_i \sigma_X(-\mathcal{A}^{\dagger} v^{(i)}).$$

We claim that $v \mapsto \sigma_X(-\mathcal{A}^{\dagger} v)$ is linear on the entirety of $\operatorname{conv} V$. To see why, note that the assumption on $v^{\star}$ relative to V means the elements of $\Phi :=$ $\{(v^{(i)}, \sigma_X(-\mathcal{A}^{\dagger} v^{(i)})) : i \in [\ell] \cup \{\star\}\}$ lie on a common hyperplane on the boundary of the epigraph $H = \{(v, t) : \sigma_X(-\mathcal{A}^{\dagger} v) \leq t\}$. Since $v \mapsto \sigma_X(-\mathcal{A}^{\dagger} v)$ is convex, H is a convex set, and there is some proper face $F \trianglelefteq H$ containing Φ. It is evident that $v \mapsto \sigma_X(-\mathcal{A}^{\dagger} v)$ is linear on the projection of that face $\hat{F} = \{v : \exists t \in \mathbb{R} \ (v, t) \in F\}$. Since $\operatorname{conv} V \subset \hat{F}$, this proves our claim regarding linearity of $v \mapsto \sigma_X(-\mathcal{A}^{\dagger} v)$ on $\operatorname{conv} V$.

By the above argument: if $v^{\star}$ is an X-circuit, then for every $\theta \in \operatorname{ri} \Delta_{\ell}$ and affinely independent $V = \{v^{(i)}\}_{i=1}^{\ell} \subset N_{\beta}$ with $\operatorname{co} V \neq \operatorname{co}\{v^{\star}\}$, we have

$$(v^{\star}, \sigma_X(-\mathcal{A}^{\dagger} v^{\star})) \neq \sum_{i=1}^{\ell} \theta_i \left(v^{(i)}, \sigma_X(-\mathcal{A}^{\dagger} v^{(i)}) \right).$$

From Carathéodory's Theorem, restricting to affinely independent $V \subset T$ is sufficient to test extremality in $\operatorname{co} T$. Therefore, every circuit $v^{\star} \in N_{\beta}$ induces an edge generator for $\operatorname{co} T$. $\qquad\square$

When considering the set "T" in Theorem 5.3.6, it is natural to expect that for polyhedral X there are only finitely many extreme rays in the cone $\operatorname{co} T$, and hence only finitely many normalized X-circuits. The remainder of this section serves to prove this fact; here we use the concept of *normal fans* from polyhedral geometry. See, e.g., [103, Chapter 7] (for the bounded case of polytopes), [139, Section 5.4] or [140, Chapter 2]. For each face F of a polyhedron P, there is an associated *outer normal cone*

$$O_P(F) = \{w : \langle z, w \rangle = \sigma_P(w) \ \forall \ z \in F\}.$$

Clearly, the support function of a polyhedron P is linear on every outer normal cone, and in particular the linear representation may be given by $\sigma_P(w) = \langle z, w \rangle$ for any $z \in F$. We obtain the *outer normal fan* of P by collecting all outer normal cones:

$$\mathcal{O}(P) = \{O_P(F) : F \trianglelefteq P\}.$$

The support of $\mathcal{O}(P)$ is the polar $(\operatorname{rec} P)^{\circ}$. The full-dimensional linearity domains of the support function are the outer normal cones of the vertices of P (see also [141, Section 1]).

Theorem 5.3.7. *If* X *is polyhedral, then* $v \in N_\beta \setminus \{0\}$ *is an* X*-circuit if and only if* $\text{co}\{v\}$ *is a ray in* $\mathcal{O}(-\mathcal{A}X + N_\beta^\circ)$. *Consequently, polyhedral* X *have finitely many normalized circuits.*

Proof. Let $P = -\mathcal{A}X + N_\beta^\circ$. Using the characterization in [65, Theorem 14.2], the polar of its recession cone can be expressed as

$$(\text{rec } P)^\circ = \{v \, : \, \sigma_X(-\mathcal{A}^\dagger v) < \infty\} \cap N_\beta,$$

where we have also used the property

$$\sigma_X(-\mathcal{A}^\dagger v) = \sup_{x \in X} \langle -\mathcal{A}^\dagger v, x \rangle = \sup_{y \in -\mathcal{A}X} \langle v, y \rangle = \sigma_{-\mathcal{A}X}(v).$$

In particular, this also gives $\sigma_X(-\mathcal{A}^\dagger v) = \sigma_P(v)$. From P construct the outer normal fan $\mathcal{O} := \mathcal{O}(P)$. We claim that $\text{co}\{v\}$ is a ray in $\mathcal{O}$.

It is clear that if a cone $K \in \mathcal{O}$ is associated to a face $F \trianglelefteq P$, then we may express $\sigma_P(v) = \langle z, v \rangle$ for any $z \in F$, and so $\sigma_P(v) \equiv \sigma_X(-\mathcal{A}^\dagger v)$ is linear on K. Since the support of $\mathcal{O}$ is $(\text{rec } P)^\circ$, the cones $K \in \mathcal{O}$ partition $(\text{rec } P)^\circ$, i.e.,

$$(\text{rec } P)^\circ = \bigcup_{K \in \mathcal{O}} \text{ri}(K),$$

and if K, K' are distinct elements in $\mathcal{O}$, then $\text{ri } K \cap \text{ri } K' = \emptyset$. Therefore, every $v \in N_\beta \setminus \{0\}$ for which $\sigma_X(-\mathcal{A}^\dagger v) < \infty$ is associated with a unique $K \in \mathcal{O}$, by way of $v \in \text{ri } K$.

Fix $v \in (\text{rec } P)^\circ$, and let K be the associated element of $\mathcal{O}$ that contains v in its relative interior. If K is of dimension greater than 1, v can be expressed as a convex combination of non-proportional $v^{(1)}, v^{(2)} \in K$ – and clearly $v \mapsto \sigma_X(-\mathcal{A}^\dagger v) \equiv \sigma_P(v)$ would be linear on the interval $[v^{(1)}, v^{(2)}]$. Thus for v to be an X-circuit, it is necessary that K be of dimension 1. Since P is a polyhedron, $\mathcal{O}$ is induced by finitely many faces. Thus there are finitely many $K \in \mathcal{O}$ with $\dim K = 1$ and in turn finitely many normalized X-circuits of $\mathcal{A}$.

Conversely, let $v^\star \in N_\beta \setminus \{0\}$ and $\text{co}\{v^\star\}$ be a ray in $\mathcal{O}$. Since $\mathcal{O}$ is supported on $(\text{rec } P)^\circ$, we have $\sigma_X(-\mathcal{A}^\dagger v) = \sigma_P(v) < \infty$.

Let $v^{(1)}, v^{(2)} \in N_\beta$ be non-proportional and $\tau \in (0, 1)$ satisfy $v^\star = \tau v^{(1)} + (1-\tau) v^{(2)}$. If $v^{(1)}$ or $v^{(2)}$ is outside of $(\text{rec } P)^\circ$, say, $v^{(1)}$, then $\sigma_X(-\mathcal{A}^\dagger v^{(1)}) = \infty$ and thus the mapping $v \mapsto \sigma_X(-\mathcal{A}^\dagger v)$ cannot be linear on $[v^{(1)}, v^{(2)}]$. Hence, we can assume that $v^{(1)}, v^{(2)} \in (\text{rec } P)^\circ$.

We have to show that the mapping

$$g : [0, 1] \to \mathbb{R}, \quad \theta \mapsto \sigma_P(\theta v^{(1)} + (1 - \theta)v^{(2)})$$

is not linear.

Consider the restriction of the fan $\mathcal{O}$ to the cone $C := \text{co}\{v^{(1)}, v^{(2)}\}$, that is, the collection of all the cones in $\{O_P(F) \cap C : F \trianglelefteq P\}$. This is a fan $\mathcal{O}'$ supported on the two-dimensional cone $S := (\text{rec } P)^\circ \cap C$. On the set S, we consider the restricted mapping $(\sigma_P)|_S : S \to \mathbb{R}, w \mapsto \sigma_P(w)$. The linearity domains of $(\sigma_P)|_S$ are the two-dimensional cones in $\mathcal{O}'$. Since $\text{co}\{v^\star\}$ is a ray in the fan $\mathcal{O}$ and thus also in the fan $\mathcal{O}'$, the vectors $v^{(1)}$ and $v^{(2)}$ are contained in different two-dimensional cones of the fan $\mathcal{O}'$. Hence, the mapping g is not linear. Altogether, this shows that $v^\star$ is an X-circuit. $\qquad\square$

Example 5.3.8. We consider as a running example the one-dimensional case of $X = [0, \infty)$ and $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\} \subset \mathbb{R}$ where we can assume $\alpha_1 < \cdots < \alpha_m$. In this running example we index by integers $i \in [m] := \{1, \ldots, m\}$ rather than by elements $\alpha \in \mathcal{A}$. Therefore we identify $\mathbb{R}^{\mathcal{A}}$ with $\mathbb{R}^m$ and use δ_i for the ith unit vector in $\mathbb{R}^m$ (for each $i \in [m]$). Under these conventions, $\mathcal{A}$ is regarded as a column vector in $\mathbb{R}^{m \times 1}$ and $\mathcal{A}^\dagger = (\alpha_1, \ldots, \alpha_m)$ is a row vector $\mathbb{R}^{1 \times m}$. We claim that the normalized X-circuits $\lambda \in \mathbb{R}^m$ are the vectors either of the form (1) $\lambda = \delta_k - \delta_j$ for $j < k$ or of the form (2)

$$\lambda = \left(\frac{\alpha_j - \alpha_i}{\alpha_k - \alpha_i}\right)\delta_k + \left(\frac{\alpha_k - \alpha_j}{\alpha_k - \alpha_i}\right)\delta_i - \delta_j \quad \text{for} \quad i < j < k.$$

Note that vectors of type (2) satisfy $\mathcal{A}^\dagger \lambda = 0$, and in fact are the unique such vectors that also satisfy $\text{supp } \lambda = \{i, j, k\}$, $\lambda_j = -1$, $\lambda_i, \lambda_k > 0$, $\langle \mathbf{1}, \lambda \rangle = 0$.

To derive this claim we consider for fixed $j \in [m]$ the polyhedron $P = -\mathcal{A}X + N_j^\circ$ from Theorem 5.3.7. It is evident that this polyhedron is a cone, that may be expressed as

$$P = \text{co}\{(-\alpha_1, \ldots, -\alpha_m)\} + \mathbb{R} \cdot \mathbf{1} - \sum_{\ell \in [m] \setminus j} \text{co}\{\delta_\ell\}.$$

The rays of its normal fan are the extreme rays of its polar

$$P^\circ = (\text{rec } P)^\circ = \{v \in \mathbb{R}^m : (-\alpha_1, \ldots, -\alpha_m) \cdot v \leq 0, \qquad (5.5)$$

$$\langle \mathbf{1}, v \rangle = 0, \ v_\ell \geq 0 \text{ for } \ell \in [m] \setminus j\}.$$

Note that this gives us exactly the set "Q" from the proof of Theorem 5.3.6. This happens because X is conic and hence the support function $\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \nu)$ evaluates to zero for every X-circuit ν. By Proposition 5.3.4, each X-circuit in N_j has at most three non-vanishing components ν_i, ν_j, ν_k, and, moreover, it has $m - 2$ of the inequalities in (5.5) binding. If all those binding inequalities are of the form $\nu_\ell \geq 0$, then with $\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \nu) < \infty$, we obtain the normalized X-circuits of $\mathcal{A}$ of type (1). Now assume that the inequality $(-\alpha_1, \ldots, -\alpha_m) \cdot \nu \leq 0$ is binding for some normalized X-circuit ν of $\mathcal{A}$. Since the sign pattern $(-, +, +)$ for (ν_i, ν_j, ν_k) in conjunction with $\langle \mathbf{1}, \nu \rangle = 0$ leads to $(-\alpha_1, \ldots, -\alpha_m) \cdot \nu < 0$, and the sign pattern $(+, +, -)$ contradicts the X-circuit condition $\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \nu) < \infty$, we obtain the normalized X-circuits of $\mathcal{A}$ of type (2).

5.4 Sublinear circuits in AGE cones

In this section, we show how the AGE cones $C_{\mathsf{X}}(\mathcal{A}, \beta)$ can be further decomposed using sublinear circuits. These decompositions lay the foundation to understand the extreme rays of the conditional SAGE cone $C_{\mathsf{X}}(\mathcal{A})$.

Our first result here is a necessary criterion for an X-AGE function f to be extremal in $C_{\mathsf{X}}(\mathcal{A}, \beta)$. The result states that any ν certifying (5.3) for f must be an X-circuit (see Theorem 5.4.2). Definition 5.4.3 introduces λ-*witnessed AGE cones* as the subset of signomials in $C_{\mathsf{X}}(\mathcal{A}, \beta)$ whose nonnegativity is certified by a given normalized vector λ. Theorem 5.4.4 then decomposes $C_{\mathsf{X}}(\mathcal{A}, \beta)$ through the λ-witnessed AGE cones, where λ is a normalized X-circuit. As a consequence, for polyhedral X, the cone $C_{\mathsf{X}}(\mathcal{A})$ is power-cone representable (see Corollary 5.4.5). The final results of this section are two elementary propositions concerning representations for primal and dual λ-witnessed AGE cones. Proposition 5.4.7 in particular is very important for a characterization of dual SAGE cones, as it reveals a multiplicative convexity property used extensively in Section 5.5.

The following lemma (proven in Subsection 5.4.1) claims it is possible to decompose an X-AGE function into simpler summands, under a local linearity condition on the support function $\nu \mapsto \sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \nu)$.

Lemma 5.4.1. *Let $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ be X-AGE with negative term $c_\beta < 0$. If ν satisfying (5.3) can be written as a convex combination $\nu = \sum_{i=1}^k \theta_i \nu^{(i)}$ of non-proportional $\nu^{(i)} \in \mathsf{N}_\beta$ and $\tilde{\nu} \mapsto \sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \tilde{\nu})$ is linear on $\mathrm{conv}\{\nu^{(i)}\}_{i=1}^k$, then f is not extremal in $C_{\mathsf{X}}(\mathcal{A}, \beta)$.*

Theorem 5.4.2. *Let $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ be X-AGE with negative term $c_\beta < 0$. If $\nu \in \mathbb{R}^{\mathcal{A}}$ satisfies (5.3) but is not an X-circuit, then f is not extremal in $C_X(\mathcal{A}, \beta)$.*

Proof. If f is an X-AGE function with $c_\beta < 0$ and ν satisfies (5.3), then we must have $\nu \neq \mathbf{0}$ and $\sigma_X(-\mathcal{A}^\dagger \nu) < \infty$. By the definition of an X-circuit, ν may be written as a convex combination $\nu = \theta \nu^{(1)} + (1 - \theta) \nu^{(2)}$ where $\bar{\nu} \mapsto \sigma_X(-\mathcal{A}^\dagger \bar{\nu})$ is linear on $[\nu^{(1)}, \nu^{(2)}]$, and furthermore the $\nu^{(i)}$ are not proportional. We can therefore invoke Lemma 5.4.1 to prove the claim. $\qquad \square$

We now eliminate the degree of freedom associated with ν laying on a ray. For each $\beta \in \mathcal{A}$, we introduce the following notation for the associated set of normalized X-circuits of $\mathcal{A}$:

$$\Lambda_X(\mathcal{A}, \beta) := \{\lambda \in \mathsf{N}_\beta \ : \ \lambda \text{ is an X-circuit of } \mathcal{A}, \ \lambda_\beta = -1\}.$$

The set of all normalized X-circuits of $\mathcal{A}$ is denoted $\Lambda_X(\mathcal{A})$. The main reason for introducing this notation is how it interacts with the following definition.

Definition 5.4.3. *Given $\lambda \in \mathsf{N}_\beta$ with $\lambda_\beta = -1$, the λ-witnessed AGE cone is*

$$C_X(\mathcal{A}, \lambda) = \left\{ \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha \ \middle| \ c_{\setminus \beta} \geq \mathbf{0} \ and \right. \tag{5.6}$$
$$\left. \prod_{\alpha \in \lambda^+} \left[\frac{c_\alpha}{\lambda_\alpha} \right]^{\lambda_\alpha} \geq -c_\beta \exp\left(\sigma_X(-\mathcal{A}^\dagger \lambda) \right) \right\}.$$

The following theorem (proven in Subsection 5.4.2) shows that every signomial in $C_X(\mathcal{A}, \lambda)$ is nonnegative on X. The term "witnessed" in "λ-witnessed AGE cone" is chosen to reflect the defining role of λ in the nonnegativity certificate. We only use λ-witnessed AGE cones for theoretical purposes, and only with $\lambda \in \Lambda_X(\mathcal{A})$. Possible computational uses are offered in Section 5.7.

Theorem 5.4.4. *The cone $C_X(\mathcal{A}, \beta)$ can be written as the convex hull of λ-witnessed AGE cones, where λ runs over the normalized X-circuits. That is,*

$$C_X(\mathcal{A}, \beta) = \operatorname{conv} \bigcup_{\lambda \in \Lambda_X(\mathcal{A}, \beta)} C_X(\mathcal{A}, \lambda).$$

To fully appreciate the significance of Theorem 5.4.4, it is necessary to consider the elementary "power cone." In our context, the primal power cone associated with a normalized vector $\lambda \in \mathsf{N}_\beta$ is

$$\operatorname{Pow}(\lambda) = \{ z \in \mathbb{R}^{\operatorname{supp} \lambda} \ : \ \textstyle\prod_{\alpha \in \lambda^+} z_\alpha^{\lambda_\alpha} \geq |z_\beta|, \ z_{\setminus \beta} \geq \mathbf{0} \};$$

the corresponding dual cone is given by

$$\text{Pow}(\lambda)^{\dagger} = \{w \in \mathbb{R}^{\text{supp}\,\lambda} \ : \ \textstyle\prod_{\alpha \in \lambda^+}[w_\alpha/\lambda_\alpha]^{\lambda_\alpha} \geq |w_\beta|, \ w_{\setminus\beta} \geq \mathbf{0}\}.$$

It should be evident that $C_{\mathsf{X}}(\mathcal{A}, \lambda)$ can be formulated in terms of a dual λ-weighted power cone; a precise formula is provided momentarily. For now we give a corollary concerning power cone representability and second-order representability of $C_{\mathsf{X}}(\mathcal{A})$ when X is a polyhedron (see [131, 142] for formal definitions).

Corollary 5.4.5. *If X is a polyhedron, then $C_{\mathsf{X}}(\mathcal{A})$ is power cone representable. If in addition $\mathcal{A}\mathsf{X}$ is rational, then $C_{\mathsf{X}}(\mathcal{A})$ is second-order representable and so its semidefinite extension degree is two.*

Corollary 5.4.5 is proven in the appendix. The first part of the corollary generalizes the case $\mathsf{X} = \mathbb{R}^n$ considered by Papp for polynomials [134]. That aspect of the corollary has uses in computational optimization when applied judiciously. The second part of Corollary 5.4.5 generalizes results by Averkov [131] and Wang and Magron [132] for ordinary SAGE polynomials, and recent results by Naumann and Theobald for several types of ordinary SAGE-like certificates [133]. We have deliberately framed the second part of the corollary in abstract terms (semidefinite extension degree), because that aspect of the corollary seems not useful for computational optimization.

We now work towards finding a simple representation of dual λ-witnessed AGE cones $C_{\mathsf{X}}(\mathcal{A}, \lambda)^{\dagger}$. We begin this process by regarding the primal as a cone of coefficients contained in $\mathbb{R}^{\mathcal{A}}$, and finding an explicit representation of the primal in terms of the elementary dual power cone $\text{Pow}(\lambda)^{\dagger}$. Towards that end we introduce a diagonal linear operator $S_\lambda : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\text{supp}\,\lambda}$ where $(S_\lambda w)_\alpha = w_\alpha$ for $\alpha \in \lambda^+$, and $(S_\lambda w)_\beta = w_\beta \exp(\sigma_{\mathsf{X}}(-\mathcal{A}^{\dagger}\lambda))$ for $\beta := \lambda^-$.

Proposition 5.4.6. *For $\lambda \in \mathsf{N}_\beta$ with $\lambda_\beta = -1$ and $\sigma_{\mathsf{X}}(-\mathcal{A}^{\dagger}\lambda) < \infty$, the λ-witnessed AGE cone admits the representation*

$$C_{\mathsf{X}}(\mathcal{A}, \lambda) = \big\{c \in \mathbb{R}^{\mathcal{A}} \,\big|\, c_{\setminus\beta} \geq \mathbf{0} \text{ and}$$
$$(S_\lambda c - r\delta_\beta) \in \text{Pow}(\lambda)^{\dagger} \text{ for some } r \geq 0\}. \qquad (5.7)$$

Proposition 5.4.6 is proven in the appendix. We can appeal to the proposition to find a representation for $C_{\mathsf{X}}(\mathcal{A}, \lambda)^{\dagger}$ which is analogous to Equation (5.6). The dual is computed by regarding the primal as a cone of *coefficients* and the derivation is given in the appendix.

Proposition 5.4.7. *For $\lambda \in \mathbb{N}_\beta$ with $\lambda_\beta = -1$ and $\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \lambda) < \infty$, the dual λ-witnessed AGE cone is given by*

$$C_{\mathsf{X}}(\mathcal{A}, \lambda)^\dagger = \left\{ v \in \mathbb{R}_+^{\mathcal{A}} \;\middle|\; \exp(\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger \lambda)) \prod_{\alpha \in \lambda^+} v_\alpha^{\lambda_\alpha} \geq v_\beta \right\}. \tag{5.8}$$

5.4.1 Proof of Lemma 5.4.1

Let c, θ, and $\{v^{(i)}\}_{i=1}^k$ be as in the lemma statement. Construct vectors $c^{(i)}$ by

$$c_\alpha^{(i)} = \begin{cases} (c_\alpha/v_\alpha)v_\alpha^{(i)} & \text{if } \alpha \in v^+ \\ 0 & \text{otherwise} \end{cases} \quad \text{for all } \alpha \in \mathcal{A} \setminus \beta, \tag{5.9}$$

and $c_\beta^{(i)} = \sigma_{\mathsf{X}}(-\mathcal{A}^\dagger v^{(i)}) + D(v_{\setminus\beta}^{(i)}, ec_{\setminus\beta}^{(i)})$. These $c^{(i)}$ define X-AGE signomials by construction, and they inherit non-proportionality from the $v^{(i)}$. We need to show that $\sum_{i=1}^k \theta_i c^{(i)} \leq c$, which will establish that f can be decomposed as a sum of these non-proportional X-AGE functions (possibly with an added posynomial).

For indices $\alpha \in v^+$, the construction (5.9) relative to v and $\{v^{(i)}\}_{i=1}^k$ actually ensures $\sum_{i=1}^k \theta_i c_\alpha^{(i)} = c_\alpha$. For indices $\alpha \in \operatorname{supp} c \setminus \operatorname{supp} v$ we have $\sum_{i=1}^k \theta_i c_\alpha^{(i)} = 0 \leq c_\alpha$. The definitions of $v^{(i)}$ ensure

$$\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger v) = \sigma_{\mathsf{X}}\left(-\mathcal{A}\left(\sum_{i=1}^k \theta_i v^{(i)}\right)\right) = \sum_{i=1}^k \theta_i \sigma_{\mathsf{X}}(-\mathcal{A}^\dagger v^{(i)}). \tag{5.10}$$

Meanwhile, (5.9) provides $v_\alpha^{(i)}/c_\alpha^{(i)} = v_\alpha/c_\alpha$, a fact we can combine this with $\sum_{i=1}^k \theta_i v_\alpha^{(i)} = v_\alpha \; \forall \; \alpha \in \mathcal{A}$ to deduce

$$\sum_{i=1}^k \theta_i D\left(v_{\setminus\beta}^{(i)}, ec_{\setminus\beta}^{(i)}\right) = D\left(v_{\setminus\beta}, ec_{\setminus\beta}\right). \tag{5.11}$$

We combine (5.10) and (5.11) to obtain the desired result

$$\sum_{i=1}^k \theta_i c_\beta^{(i)} = \sum_{i=1}^k \theta_i \left(\sigma_{\mathsf{X}}(-\mathcal{A}^\dagger v^{(i)}) + D\left(v_{\setminus\beta}^{(i)}, ec_{\setminus\beta}^{(i)}\right)\right)$$

$$= \sigma_{\mathsf{X}}(-\mathcal{A}^\dagger v) + D(v_{\setminus\beta}, ec_{\setminus\beta}) \leq c_\beta.$$

5.4.2 Proof of Theorem 5.4.4

Our proof requires the following proposition.

Proposition 5.4.8. *For fixed λ in the interior of the m-dimensional probability simplex and $c = (c_0, c_1, \ldots, c_m) \in \mathbb{R}^{m+1}$ with $(c_1, \ldots, c_m) \geq \mathbf{0}$, we have*

$$-c_0 \leq \prod_{i=1}^m [c_i/\lambda_i]^{\lambda_i} \quad \Leftrightarrow \quad \text{some } v \in \mathbb{R}_+^m \text{ satisfies } v \parallel \lambda \text{ and } D(v, c_{\setminus 0}) - \langle \mathbf{1}, v \rangle \leq c_0$$

– where $v \parallel \lambda$ means v is proportional to λ.

Proof. The claim is trivial when $c_0 \geq 0$, and so we consider $c_0 < 0$. Note that in this case, $\prod_{i=1}^{m} [c_i/\lambda_i]^{\lambda_i}$ must be positive, and $D(v, c_{\backslash 0})$ must be finite: both of these conditions occur precisely when $c_i > 0$ for all $1 \leq i \leq m$. We therefore can rewrite $-c_0 = |c_0| \leq \prod_{i=1}^{m} [c_i/\lambda_i]^{\lambda_i}$ as $1 \leq \prod_{i=1}^{m} [c_i/(|c_0|\lambda_i)]^{\lambda_i}$, and by taking the log of both sides, obtain $D(v, c_{\backslash 0}) - \langle \mathbf{1}, v \rangle \leq c_0$ for $v = |c_0|\lambda$. For the other direction, one may write the proportionality relationship $v \parallel \lambda$ as $v = s\lambda$, and minimize $D(s\lambda, c_{\backslash 0}) - s$ over $s \geq 0$ to obtain $-\prod_{i=1}^{m} [c_i/\lambda_i]^{\lambda_i}$. $\qquad\square$

We now turn to proving Theorem 5.4.4. First, observe that Theorem 5.4.2 tells us $C_X(\mathcal{A}, \beta)$ may be expressed as the convex hull of X-AGE functions $f = \sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ where (c, v) satisfies (5.3) for some X-circuit v. Therefore it suffices to show that (i) for any such function, the normalized X-circuit $\lambda = v/|v_\beta|$ is such that (c, λ) satisfy the condition in (5.6), and (ii) if *any* (c, λ) satisfy (5.6), then the resulting signomial is nonnegative on X. We will actually do both of these in one step.

Suppose $v \in N_\beta$ is restricted to satisfy $v = s\lambda$ for a variable $s \geq 0$ and a fixed $\lambda \in \Lambda_X(\mathcal{A}, \beta)$. It suffices to show that the set of $c \in \mathbb{R}^{\mathcal{A}}$ for which

$$\exists s \geq 0 \, : \, v = s\lambda \text{ and } \sigma_X(-\mathcal{A}^\dagger v) + D(v_{\backslash \beta}, ec_{\backslash \beta}) \leq c_\beta$$

is the same as (5.6).

Let $r(v) = \sigma_X(-\mathcal{A}^\dagger v) + D(v_{\backslash \beta}, ec_{\backslash \beta})$. Apply positive homogeneity of the support function to see $\sigma_X(-\mathcal{A}^\dagger v) = |v_\beta|\sigma_X(\mathcal{A}^\dagger v/|v_\beta|)$, and use $v = s\lambda$ to infer $s = |v_\beta|$ and $\sigma_X(-\mathcal{A}^\dagger v/|v_\beta|) = \sigma_X(-\mathcal{A}^\dagger \lambda)$. Abbreviate $d := \sigma_X(-\mathcal{A}^\dagger \lambda)$ and substitute $\sum_{\alpha \in \lambda^+} v_\alpha = |v_\beta|$ to obtain

$$r(v) = \sum_{\alpha \in \lambda^+} (v_\alpha \log(v_\alpha/c_\alpha) - v_\alpha + v_\alpha d) \,.$$

The term d can be moved into the logarithm by $v_\alpha d = v_\alpha \log(1/\exp(-d))$. For $\alpha \in \lambda^+$ we define scaled terms $\tilde{c}_\alpha = c_\alpha \exp(-d)$, so that

$$r(v) = \sum_{\alpha \in \lambda^+} v_\alpha \log(v_\alpha/\tilde{c}_\alpha) - v_\alpha.$$

By Proposition 5.4.8, there exists a $v = s\lambda$ for which $r(v) \leq c_\beta$ if and only if

$$-c_\beta \leq \prod_{\alpha \in \lambda^+} [\tilde{c}_\alpha/\lambda_\alpha]^{\lambda_\alpha}. \tag{5.12}$$

Since $[\tilde{c}_\alpha/\lambda_\alpha]^{\lambda_\alpha} = [c_\alpha/\lambda_\alpha]^{\lambda_\alpha} (\exp(-d))^{\lambda_\alpha}$ and $\prod_{\alpha \in \lambda^+} (\exp(-d))^{\lambda_\alpha} = \exp(-d)$, (5.12) can be recognized as the inequality occurring within (5.6), which completes the proof.

5.5 Reduced sublinear circuits in SAGE cones

The previous section showed that an X-SAGE cone is generated by X-circuits. Here we seek a much sharper characterization: are all X-circuits really necessary? The answer to this question depends on whether one means to reconstruct an individual AGE cone, or the larger SAGE cone. For example, by reinterpreting results from [59], we may infer that every simplicial $\mathbb{R}^n$-circuit $\lambda \in \Lambda_{\mathbb{R}^n}(\mathcal{A}, \beta)$ generates a cone $C_{\mathbb{R}^n}(\mathcal{A}, \lambda)$ that contains an extreme ray of $C_{\mathbb{R}^n}(\mathcal{A}, \beta)$. In this way, every $\mathbb{R}^n$-circuit is needed if one requires complete reconstruction of individual AGE cones. However, Katthän, Naumann, and Theobald showed that many extreme rays of AGE cones are *not* extreme when considered in the sum $C_{\mathbb{R}^n}(\mathcal{A}) = \sum_{\beta \in \mathcal{A}} C_{\mathbb{R}^n}(\mathcal{A}, \beta)$. Specifically, an $\mathbb{R}^n$-circuit $\lambda \in \Lambda_{\mathbb{R}^n}(\mathcal{A})$ is only needed in $C_{\mathbb{R}^n}(\mathcal{A})$ if exactly one element of $\mathcal{A}$ hits the relative interior of $\text{conv}(\text{supp}\,\lambda)$ [130, Proposition 4.4]. Circuits satisfying this property were called *reduced*. The goal of this section is to develop a reducedness criterion for X-circuits that yields the most efficient construction of $C_X(\mathcal{A})$ by λ-witnessed AGE cones, see Theorems 5.5.5 and 5.5.6. Achieving this goal is more difficult than obtaining the results from earlier sections. Therefore we begin by summarizing and discussing the results, and we provide proofs in later subsections.

5.5.1 Definitions, results, and discussion

The definition of a reduced $\mathbb{R}^n$-circuit is of a purely combinatorial nature, involving the circuit's support. This is appropriate because when speaking of affine-linear simplicial circuits, the normalized vector representation λ is completely determined by its support. In the context of X-circuits, we no longer have this property. Therefore when developing reduced X-*circuits* it is useful to have a different characterization of reduced $\mathbb{R}^n$-*circuits*. Here we can consider how Forsgård and de Wolff defined the *Reznick cone of $\mathcal{A}$* as the conic hull $\text{co}\,\Lambda_{\mathbb{R}^n}(\mathcal{A})$ and – in the language of Katthän et al. – subsequently proved that an $\mathbb{R}^n$-circuit λ is an edge generator of the Reznick cone if and only if it is reduced [21].

Our definition of reduced X-circuits involves edge generators of a certain cone in one higher dimension than the Reznick cone. To describe the cone and facilitate later analysis, we need the following definition.

Definition 5.5.1. *The functional form of an* X-*circuit* $\nu \in \mathbb{R}^{\mathcal{A}}$ *is* $\phi_\nu : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}$ *defined by*

$$\phi_\nu(y) = \sum_{\alpha \in \mathcal{A}} y_\alpha \nu_\alpha + \sigma_X(-\mathcal{A}^\dagger \nu).$$

We routinely overload notation and use $\phi_\nu = (\nu, \sigma_X(-\mathcal{A}^\dagger\nu)) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R}$ to denote the functional form of a given X-circuit. When representing the functional form of an X-circuit by a vector in $\mathbb{R}^{\mathcal{A}} \times \mathbb{R}$, the scalar $\phi_\nu(y)$ can be expressed as an inner product $\phi_\nu(y) = (y, 1)^\top \phi_\nu$.

Definition 5.5.2. *The circuit graph of* $(\mathcal{A}, X)$ *is*

$$G_X(\mathcal{A}) = \text{co}\left(\{\phi_\lambda \; : \; \lambda \in \Lambda_X(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}\right),$$

where $(\mathbf{0}, 1) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R}$.

The idea of generating a cone from augmented circuit vectors $(\nu, \sigma_X(-\mathcal{A}^\dagger\nu)) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R}$ clearly parallels Theorem 5.3.6. While the cones from Theorem 5.3.6 are considered for one $\beta \in \mathcal{A}$ at a time, the circuit graph accounts for all X-circuits at once. The circuit graph also includes an extra generator that ultimately serves to make the following definition more stringent.

Definition 5.5.3. *The reduced X-circuits of* $\mathcal{A}$ *are the vectors* ν *where* $\nu/\|\nu\|_\infty \in \Lambda_X(\mathcal{A})$ *and the corresponding functional form* ϕ_ν *generates an extreme ray of* $G_X(\mathcal{A})$. *The set of normalized reduced X-circuits is henceforth denoted* $\Lambda_X^\star(\mathcal{A})$.

There is a subtle issue here that in order for reduced X-circuits to be of any use to us, the circuit graph must be pointed (else $G_X(\mathcal{A})$ would have no extreme rays whatsoever). We show later in this section that our stated assumption of linear independence of $\{e^\alpha\}_{\alpha \in \mathcal{A}}$ on X ensures $G_X(\mathcal{A})$ is pointed. Regardless of whether or not the circuit graph is pointed, we have the following theorem.

Theorem 5.5.4. $C_X(\mathcal{A})^\dagger = \text{cl}\{\exp y \; : \; (y, 1) \in G_X(\mathcal{A})^\dagger\}$.

Theorem 5.5.4 is noteworthy in several respects. It demonstrates that $C_X(\mathcal{A})^\dagger$ is convex in the usual sense and convex under a logarithmic transformation $S \mapsto \log S = \{y : \exp y \in S\}$. This *multiplicative convexity* is a significant structural property. For example, if we know that the log of the moment cone $\text{cl}(\text{co}\{\exp(\mathcal{A}x) : x \in X\})$ is not convex, then it should be that $C_X(\mathcal{A})$ does not contain all X-nonnegative signomials supported on $\mathcal{A}$. Additionally, Theorem 5.5.4 can be reverse-engineered to arrive at the concept of a reduced X-circuit: the definition is chosen so that $(y, 1)$ belongs to $G_X(\mathcal{A})^\dagger$ if and only if $\phi_\lambda(y) \geq 0$ for all λ in $\Lambda_X^\star(\mathcal{A})$. Here, Theorem 5.5.4 is a tool that we combine with convex duality to obtain the following results.

Theorem 5.5.5. *If* $\Lambda_X(\mathcal{A})$ *is empty, then* $C_X(\mathcal{A}) = \mathbb{R}_+^{\mathcal{A}}$. *Otherwise,*

$$C_X(\mathcal{A}) = \mathrm{cl}\left(\mathrm{conv}\bigcup\left\{C_X(\mathcal{A}, \lambda) \,:\, \lambda \in \Lambda_X^{\star}(\mathcal{A})\right\}\right). \tag{5.13}$$

We point out how Theorem 5.5.5 involves a closure around the union over λ-witnessed AGE cones, while Theorem 5.4.4 has no such closure. The need for the closure here stems from an application of an infinite version of conic duality in the course of the theorem's proof, while our proof of Theorem 5.4.4 required no duality at all. The requisite use of conic duality is simpler when X is a polyhedron, as the following theorem suggests.

Theorem 5.5.6. *If* X *is a polyhedron and* $\Lambda_X(\mathcal{A})$ *is nonempty, then the associated conditional SAGE cone is given by the finite Minkowski sum*

$$C_X(\mathcal{A}) = \sum_{\lambda \in \Lambda_X^{\star}(\mathcal{A})} C_X(\mathcal{A}, \lambda). \tag{5.14}$$

Moreover, there is no $\Lambda \subsetneq \Lambda_X^{\star}(\mathcal{A})$ *for which* $C_X(\mathcal{A}) = \sum_{\lambda \in \Lambda} C_X(\mathcal{A}, \lambda)$.

The first part of Theorem 5.5.6 follows easily from the arguments we use to prove Theorem 5.5.5. The second part of the theorem is much more delicate, and in fact is the reason why $G_X(\mathcal{A})$ is defined in the manner of 5.5.2, rather than merely $\mathrm{co}\{\phi_\lambda \,:\, \lambda \in \Lambda_X(\mathcal{A})\}$.

The task of actually finding the reduced X-circuits of $\mathcal{A}$ is difficult. When X is a polyhedron there are finitely many such X-circuits, but the naive method for finding them involves Fourier-Motzkin elimination on a set of potentially very high dimension. There is more hope for this problem when X is a cone. In that case, X-circuits are the extreme rays of $(\ker(\mathcal{A}^\dagger) + \mathcal{A}^{-1}X^\dagger) \cap N_\beta$ for $\beta \in \mathcal{A}$,[2] and no lifting is needed to find these extreme rays with a computer. The reduced X-circuits could then be computed by finding the extreme rays of the convex cone generated by the X-circuits. The following detailed example finds the reduced X-circuits of $\mathcal{A}$ in the univariate case with the cone $X = [0, \infty)$. The claim made in the example is used in Section 5.6.

Example 5.5.7. We continue the running example of $X = [0, \infty)$ from Example 5.3.8. In particular recall $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$ for $\alpha_1 < \cdots < \alpha_m$, indexing by $i \in [m]$, and working with standard basis $\delta_i \in \mathbb{R}^m$. We claim that

$$\Lambda_{[0,\infty)}^{\star}(\mathcal{A}) = \{\delta_2 - \delta_1\} \cup \Lambda_{\mathbb{R}}^{\star}(\mathcal{A}) \tag{5.15}$$

---

[2]Recall, $\mathcal{A}^{-1}$ is the pseudo-inverse of $\mathcal{A} : \mathbb{R}^n \to \mathbb{R}^{\mathcal{A}}$.

where we have the following formula from [130, Prop. 4.4]

$$\Lambda_{\mathbb{R}}^{\star}(\mathcal{A}) = \left\{ \left(\frac{\alpha_{i+1} - \alpha_i}{\alpha_{i+1} - \alpha_{i-1}} \right) \delta_{i-1} + \left(\frac{\alpha_i - \alpha_{i-1}}{\alpha_{i+1} - \alpha_{i-1}} \right) \delta_{i+1} - \delta_i \ : \ 1 < i < m \right\}.$$

As a first step towards seeing this, observe that since $X = [0, \infty)$ is a cone, the functional form of a $[0, \infty)$-circuit ν is simply $\phi_\nu(y) = \sum_{i=1}^{m} y_i \nu_i$. Hence, the reduced $[0, \infty)$-circuits are exactly the edge generators of the cone $\operatorname{co} \Lambda_{[0,\infty)}$ generated by all the $[0, \infty)$-circuits of types (1) and (2) listed in Example 5.3.8. Therefore, we have to show that $\{\delta_2 - \delta_1\} \cup \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$ are exactly the normalized edge generators of $\operatorname{co} \Lambda_{[0,\infty)}$.

For the X-circuits $\delta_j - \delta_i$ $(j > i)$ of type (1) in Example 5.3.8, we show they decompose if $j > i + 1$ or $i > 1$. For $j > i + 1$, this is apparent from the decomposition

$$\delta_j - \delta_i = (\delta_j - \delta_{j-1}) + (\delta_{j-1} - \delta_i).$$

For $j = i + 1$ and $i > 1$, we can use the decomposition

$$\delta_{i+1} - \delta_i = \left(-\frac{\alpha_{i+1} - \alpha_i}{\alpha_i - \alpha_{i-1}} \delta_{i-1} + \frac{\alpha_{i+1} - \alpha_i}{\alpha_i - \alpha_{i-1}} \delta_i \right) + \left(\frac{\alpha_{i+1} - \alpha_i}{\alpha_i - \alpha_{i-1}} \delta_{i-1} - \frac{\alpha_{i+1} - \alpha_{i-1}}{\alpha_i - \alpha_{i-1}} \delta_i + \delta_{i+1} \right)$$

into X-circuits with three non-vanishing components. As final consideration for type (1), the X-circuit $\delta_2 - \delta_1$ cannot be written as a conic combination of X-circuits with three non-zero entries, because any conic combination of those X-circuits has a positive entry in its non-vanishing component with maximal index. For X-circuits of type (2) from Example 5.3.8, simply note that these are also $\mathbb{R}$-circuits. Therefore a necessary condition for a type (2) X-circuit λ to be extremal in $\operatorname{co} \Lambda_{[0,\infty)}$ is that λ belongs to $\Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$.

It remains to show that none of the remaining X-circuits can be written as a convex combination of the others. First note that an X-circuit $\nu \in \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$ cannot be decomposed into a sum which involves an X-circuit $\tilde{\nu}$ with two vanishing components. Namely, since $\mathcal{A}^{\dagger} \nu = 0$ and $\mathcal{A} \tilde{\nu} > 0$, we would obtain for the other summand $\nu - \tilde{\nu}$ the property $\mathcal{A}^{\dagger}(\nu - \tilde{\nu}) < 0$ and thus $\sigma_{[0,\infty)}(-\mathcal{A}^{\dagger}(\nu - \tilde{\nu})) = \infty$, a contradiction. And of course it is trivially true that no element $\lambda \in \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$ can be written as a convex combination of other such elements. Since $\operatorname{co} \Lambda_{[0,\infty)}$ is finitely generated and there is no $S \subsetneq \{\delta_2 - \delta_1\} \cup \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$ for which $\operatorname{co} \Lambda_{[0,\infty)} = \operatorname{co} S$, we conclude that $\{\delta_2 - \delta_1\} \cup \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$ are the reduced X-circuits of $\mathcal{A}$.

The remainder of this section is organized as follows. Subsection 5.5.2 proves Theorem 5.5.4, which is instrumental in later subsections. In Subsection 5.5.3 we

introduce and prove a certain representation result for the circuit graph. Given the groundwork laid in these two subsections, Subsection 5.5.4 proves Theorem 5.5.5 in very short order. Subsection 5.5.5 proves Theorem 5.5.6 by refining the arguments from Subsection 5.5.4.

5.5.2   Proof of Theorem 5.5.4

Proof of Theorem 5.5.4. Use Rockafellar's [65, Corollary 16.5.2] to invoke Theorem 5.4.4 from a dual point of view, which gives $C_X(\mathcal{A}, \beta)^\dagger = \bigcap C_X(\mathcal{A}, \lambda)^\dagger$, where the intersection runs over all $\lambda \in \Lambda_X(\mathcal{A}, \beta)$. Then Proposition 5.4.7 implies

$$C_X(\mathcal{A})^\dagger = \left\{ v \in \mathbb{R}_+^{\mathcal{A}} \;\middle|\; \text{for all } \lambda \text{ in } \Lambda_X(\mathcal{A}) \text{ and } \beta := \lambda^-, \right.$$
$$\left. \text{we have}\ \ \exp(\sigma_X(-\mathcal{A}^\dagger \lambda)) \prod_{\alpha \in \lambda^+} v_\alpha^{\lambda_\alpha} \geq v_\beta \right\}. \qquad (5.16)$$

We claim that $C_X(\mathcal{A})^\dagger$ can be represented as the closure of its intersection with the positive orthant, that is, $C_X(\mathcal{A})^\dagger = \mathrm{cl}\left(C_X(\mathcal{A})^\dagger \cap \mathbb{R}_{++}^{\mathcal{A}}\right)$. Since $C_X(\mathcal{A})$ contains all posynomials and is contained in the nonnegativity cone, the dual $C_X(\mathcal{A})^\dagger$ contains the moment cone, but is still contained in the nonnegative orthant. As we have assumed X is nonempty, $C_X(\mathcal{A})^\dagger$ must contain a point $\exp(\mathcal{A}x) \in \mathbb{R}_{++}^{\mathcal{A}}$, so $C_X(\mathcal{A})^\dagger \cap$ ri $\mathbb{R}_+^{\mathcal{A}} \neq \emptyset$. We recall the old Proposition 2.6.4 from the start of this thesis. It tells us that $C_X(\mathcal{A})^\dagger = \mathrm{cl}\left(C_X(\mathcal{A})^\dagger \cap \mathbb{R}_{++}^{\mathcal{A}}\right)$.

When considering $C_X(\mathcal{A})^\dagger$ only over the positive orthant, the inequalities

$$\exp(\sigma_X(-\mathcal{A}^\dagger \lambda)) \prod_{\alpha \in \lambda^+} v_\alpha^{\lambda_\alpha} \geq v_\beta$$

appearing in (5.16) may be rewritten as

$$\textstyle\sum_{\alpha \in \lambda^+} \lambda_\alpha \log v_\alpha - \log v_\beta + \sigma_X(-\mathcal{A}^\dagger \lambda) \equiv \phi_\lambda(y) \geq 0,$$

where we used $\lambda_\beta = -1$ and $y := \log v \in \mathbb{R}^{\mathcal{A}}$. Hence,

$$\begin{aligned}
C_X(\mathcal{A})^\dagger &= \mathrm{cl}\{\exp y \;:\; \phi_\lambda(y) \geq 0 \;\; \forall \lambda \in \Lambda_X(\mathcal{A})\} \\
&= \mathrm{cl}\{\exp y \;:\; (y, 1)^\top (\lambda, \tau) \geq 0 \;\; \forall \lambda \in \Lambda_X(\mathcal{A}), \; \tau \geq \sigma_X(-\mathcal{A}^\dagger \lambda)\} \\
&= \mathrm{cl}\{\exp y \;:\; (y, 1)^\top (\nu, \tau) \geq 0 \;\; \forall (\nu, \tau) \in G_X(\mathcal{A})\}.
\end{aligned}$$

By the definition of the dual cone from convex analysis, the property $(y, 1)^\top (\nu, \tau) \geq 0 \; \forall (\nu, \tau) \in G_X(\mathcal{A})$ is the same as $(y, 1) \in G_X(\mathcal{A})^\dagger$. This completes the proof. $\square$

The ability to represent $C_X(\mathcal{A})^\dagger$ in terms of $G_X(\mathcal{A})^\dagger$ is key to our proofs of Theorems 5.5.5 and 5.5.6. Note that the theorem remains true when $G_X(\mathcal{A})$ is replaced by the smaller set $\mathrm{co}\{\phi_\lambda : \lambda \in \Lambda_X(\mathcal{A})\}$, because the term $(\mathbf{0}, 1)$ simply requires $(\mathbf{y}, t) \in G_X(\mathcal{A})^\dagger$ to have $t \geq 0$.

5.5.3 Topological properties of the circuit graph

Theorem 5.5.8. $G_X(\mathcal{A}) = \mathrm{co}\left(\{\phi_\lambda : \lambda \in \Lambda_X^\star(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}\right).$

The proof of this theorem essentially reduces to showing that $G_X(\mathcal{A})$ is pointed and closed. The pointedness of the circuit graph is easy to show, but closedness is a more delicate matter. In fact – our proof that $G_X(\mathcal{A})$ is closed relies on the fact that it is pointed. We therefore prove pointedness before discussing closedness any further.

Lemma 5.5.9. *The closure of the circuit graph contains no lines.*

Proof. We focus on proving $G_X(\mathcal{A})^\dagger$ is full-dimensional. Let $|\mathcal{A}| = m$. We assumed at the start of this chapter that the moment cone $M_X(\mathcal{A}) := \mathrm{co}\{\exp(\mathcal{A}\mathbf{x}) : \mathbf{x} \in X\}$ was full-dimensional, i.e., $\dim M_X(\mathcal{A}) = m$; we use that assumption here. Specifically, since $C_X(\mathcal{A})$ is contained within the nonnegativity cone, we have that $M_X(\mathcal{A}) \subset C_X(\mathcal{A})^\dagger$ and so $\dim C_X(\mathcal{A})^\dagger = m$. By Theorem 5.5.4 and continuity of the exponential function, we see that if $\dim C_X(\mathcal{A})^\dagger = m$, then the preimage $S := \{\mathbf{y} : (\mathbf{y}, 1) \in G_X(\mathcal{A})^\dagger\}$ likewise has dimension m. Consider the induced cone associated with S:

$$\mathrm{indco}(S) = \mathrm{cl}\{(\mathbf{y}, t) : t > 0, \mathbf{y}/t \in S\} = \mathrm{cl}\{(\mathbf{y}, t) : t > 0, (\mathbf{y}, t) \in G_X(\mathcal{A})^\dagger\}.$$

The rightmost expression in the above display tells us $\mathrm{indco}(S) \subset G_X(\mathcal{A})^\dagger$. We claim without proof that since S is a full-dimensional convex set, $\mathrm{indco}(S)$ is similarly full-dimensional. Taking this claim as given, $\mathrm{indco}(S) \subset G_X(\mathcal{A})^\dagger$ implies $G_X(\mathcal{A})^\dagger$ is full-dimensional. Because $G_X(\mathcal{A})^\dagger$ is full-dimensional, $\mathrm{cl}\, G_X(\mathcal{A}) = G_X(\mathcal{A})^{\dagger\dagger} \supset G_X(\mathcal{A})$ contains no lines. $\square$

In the special case where X is a polyhedron, closedness of $G_X(\mathcal{A})$ follows from Theorem 5.3.7, which tells us that $\Lambda_X(\mathcal{A})$ is finite. To prove closedness for arbitrary convex sets X we need to more carefully appeal to properties of the generating set $\{\phi_\lambda : \lambda \in \Lambda_X(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}$.

Lemma 5.5.10. *The circuit graph is closed.*

Proof. Let $S_\beta = \{(\lambda, \sigma_X(-\mathcal{A}^\dagger\lambda)) : \lambda \in \Lambda_X(\mathcal{A}, \beta)\}$. By Theorem 5.3.6, the elements $\phi_\lambda \in S_\beta$ are edge generators for the closed convex cone

$$T_\beta = \mathrm{co}\{(\nu, \sigma_X(-\mathcal{A}^\dagger\nu)) : \nu \in N_\beta, \sigma_X(-\mathcal{A}^\dagger\nu) < \infty\}.$$

From S_β we form $S'_\beta := \mathrm{conv}\, S_\beta$, and find S'_β is isomorphic to

$$S'_\beta = \{\phi_\lambda \in T_\beta : \lambda_\beta = -1\}.$$

Because S_β is bounded, S'_β is likewise bounded. Because S'_β is a slice of a closed convex cone T_β, we have that S'_β is closed. Therefore we conclude S'_β is compact.

Now define $S' = (\bigcup_{\beta\in\mathcal{A}} S'_\beta) \cup \{(\mathbf{0}, 1)\}$. The set S' is a compact generating set for $G_X(\mathcal{A})$ which does not contain the origin. Since $\mathrm{cl}\, G_X(\mathcal{A})$ is known to contain no lines (Lemma 5.5.9), we apply Proposition B.1.1 to S', $\mathrm{co}\, S'$ to infer that $\mathrm{co}\, S' = G_X(\mathcal{A})$ is closed. $\qquad\square$

Proof of Theorem 5.5.8. Lemmas 5.5.9 and 5.5.10 show $G_X(\mathcal{A})$ is closed and pointed. By [65, Corollary 18.5.2], we have that $G_X(\mathcal{A})$ may be expressed as the conic hull of any set of vectors containing all of its extreme rays. Since $S = \{\phi_\lambda : \lambda \in \Lambda_X(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}$ is a generating set for $G_X(\mathcal{A})$, it must contain all extreme rays of $G_X(\mathcal{A})$. However, by definition of $\Lambda_X^\star(\mathcal{A})$, if λ does not belong to $\Lambda_X^\star(\mathcal{A})$, then $\phi_\lambda \in S$ does not generate an extreme ray of $G_X(\mathcal{A})$. We may therefore form $T = S \setminus \{\phi_\lambda : \lambda \notin \Lambda_X^\star(\mathcal{A})\}$ and still find $G_X(\mathcal{A}) = \mathrm{co}\, T$. This proves the theorem. $\qquad\square$

5.5.4 Proof of Theorem 5.5.5.

Proof of Theorem 5.5.5. Using the representation

$$G_X(\mathcal{A}) = \mathrm{co}\left(\{\phi_\lambda : \lambda \in \Lambda_X^\star(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}\right)$$

provided by Theorem 5.5.8, we can express

$$(\mathbf{y}, 1) \in G_X(\mathcal{A})^\dagger \Leftrightarrow (\mathbf{y}, 1)^\top(\lambda, \sigma_X(-\mathcal{A}^\dagger\lambda)) \geq 0 \;\forall\; \lambda \in \Lambda_X^\star(\mathcal{A}). \tag{5.17}$$

We obtain the following refinement of Equation (5.16), by combining (5.17) with Theorem 5.5.4:

$$C_X(\mathcal{A})^\dagger = \left\{\mathbf{v} \in \mathbb{R}_+^{\mathcal{A}} : \exp(\sigma_X(-\mathcal{A}^\dagger\lambda)) \prod_{\alpha\in\lambda^+} v_\alpha^{\lambda_\alpha} \geq v_\beta \right. \tag{5.18}$$

$$\left. \text{for every } \lambda \in \Lambda_X^\star(\mathcal{A}) \text{ and } \beta := \lambda^- \right\}.$$

Of course, Equation (5.18) can be written as $C_X(\mathcal{A})^\dagger = \bigcap_{\lambda\in\Lambda_X^\star(\mathcal{A})} C_X(\mathcal{A}, \lambda)^\dagger$. We appeal to conic duality principles (again, [65, Corollary 16.5.2]) to obtain the claim of the theorem. $\qquad\square$

5.5.5 Proof of Theorem 5.5.6

A conceptual message from the last section is that it can be very useful to analyze $C_X(\mathcal{A})$ in terms of the vectors y where $\exp y$ belongs to $C_X(\mathcal{A})^\dagger$. This section will hammer that message home. We begin with the lemma that ultimately led us to define $G_X(\mathcal{A})$ as per Definition 5.5.2, rather than as the simpler set $\text{co}\{\phi_\lambda \,:\, \lambda \in \Lambda_X(\mathcal{A})\}$.

Lemma 5.5.11. *If* X *is polyhedral and* $\Lambda \subsetneq \Lambda_X^\star(\mathcal{A})$*, then there must exist a* $\tilde{y} \in \mathbb{R}^{\mathcal{A}}$ *satisfying* $\phi_{\lambda'}(\tilde{y}) \geq 0$ *for all* $\lambda' \in \Lambda$*, yet for some* $\lambda \in \Lambda_X^\star(\mathcal{A}) \setminus \Lambda$ *we have* $\phi_\lambda(\tilde{y}) < 0$*.*

Proof. Let $T_1 = \{\phi_\lambda \,:\, \lambda \in \Lambda_X^\star(\mathcal{A})\} \cup \{(\mathbf{0}, 1)\}$ and $T_2 = \{\phi_\lambda \,:\, \lambda \in \Lambda\} \cup \{(\mathbf{0}, 1)\}$. Of course, a vector $\tilde{y}$ satisfies $\phi_{\lambda'}(\tilde{y}) \geq 0$ for all $\lambda' \in \Lambda$ if and only if $(\tilde{y}, 1) \in (\text{co } T_2)^\dagger$. We will show that given the polyhedrality of X and the assumption on Λ, there exists a vector $\tilde{y}$ for which $(\tilde{y}, 1) \in (\text{co } T_2)^\dagger \setminus (\text{co } T_1)^\dagger$. The result will follow since membership of vectors $(y, 1) \in (\text{co } T_1)^\dagger$ is equivalent to $\phi_\lambda(y) \geq 0$ for all $\lambda \in \Lambda_X^\star(\mathcal{A})$.

Since X is polyhedral, Theorem 5.3.7 tells us $\Lambda_X(\mathcal{A})$ is finite, so $\Lambda_X(\mathcal{A})$ is closed and $\Lambda_X^\star(\mathcal{A})$ is finite. From closedness of $\Lambda_X(\mathcal{A})$ we have $G_X(\mathcal{A}) = \text{co } T_1$, and in particular every $\phi_\lambda \in T_1 \setminus \{(\mathbf{0}, 1)\}$ is known to generate an extreme ray in $G_X(\mathcal{A})$. Since $\Lambda \subsetneq \Lambda_X^\star(\mathcal{A})$, there exists a $\phi_\lambda \in T_1 \setminus T_2$ which generates an extreme ray of $G_X(\mathcal{A})$. Therefore $\text{co } T_2$ is a strict subset of $\text{co } T_1 \equiv G_X(\mathcal{A})$. We may take dual cones to find $(\text{co } T_2)^\dagger \supsetneq (\text{co } T_1)^\dagger$. Note that since T_1 and T_2 contain $\{(\mathbf{0}, 1)\}$, the dual cones must be contained in $K = \mathbb{R}^{\mathcal{A}} \times \mathbb{R}_+$. Furthermore, since X is presumed nonempty, Theorem 5.5.4 tells us there exists a point $(y, 1) \in (\text{co } T_1)^\dagger$, so the relative interiors of $(\text{co } T_1)^\dagger$ and $(\text{co } T_2)^\dagger$ are contained within the relative interior of K. As our last step, use the fact that if one closed polyhedral cone strictly contains another closed polyhedral cone, then there exists a point in the relative interior of the larger cone which may be separated from the smaller cone; apply this to $(\text{co } T_2)^\dagger \supsetneq (\text{co } T_1)^\dagger$ to find a point $(y', t') \in \text{ri}((\text{co } T_2)^\dagger) \setminus (\text{co } T_1)^\dagger$ with $t' > 0$. From this (y', t') we rescale $\tilde{y} = y'/t'$ so that $(\tilde{y}, 1) \in (\text{co } T_2)^\dagger \setminus (\text{co } T_1)^\dagger$. $\qquad \square$

Our next lemma shows how to take a condition stated in terms of Lemma 5.5.11, and deduce a statement about $C_X(\mathcal{A})^\dagger$. The lemma's proof requires only that X be nonempty and convex.

Lemma 5.5.12. *If* $\tilde{y} \in \mathbb{R}^{\mathcal{A}}$ *satisfies* $\phi_\lambda(\tilde{y}) < 0$ *for some* $\lambda \in \Lambda_X(\mathcal{A})$*, then* $\exp \tilde{y} \notin C_X(\mathcal{A})^\dagger$*.*

Proof. In this proof we use $a^\top b$ to take the dot product of vectors a, b, rather than using the notation of inner products. We will find a vector $z \in \mathbb{R}^{\mathcal{A}}$ where $0 \leq z^\top \exp y$ for all $\exp y \in C_X(\mathcal{A})^\dagger$, and yet $z^\top \exp \tilde{y} < 0$. By continuity, the condition that $0 \leq z^\top \exp y$ for all $\exp y \in C_X(\mathcal{A})^\dagger$ will imply the slightly stronger statement that $0 \leq z^\top v$ for all $v \in C_X(\mathcal{A})^\dagger$. Therefore z will evidently serve as a separating hyperplane to prove the desired claim. Let $\beta := \lambda^-$.

Since $\lambda \in \Lambda_X(\mathcal{A})$, Theorem 5.5.4 says that $\phi_\lambda(y) \geq 0$ whenever $\exp y \in C_X(\mathcal{A})^\dagger$. Combine $\phi_\lambda(\tilde{y}) < 0$ with strict monotonicity of the exponential function to conclude

$$\exp(\phi_\lambda(\tilde{y})) < 1 \leq \exp(\phi_\lambda(y)) \quad \text{for all} \quad \exp y \in C_X(\mathcal{A})^\dagger. \tag{5.19}$$

Notice that taking a difference

$$\phi_\lambda(y) - \phi_\lambda(\tilde{y}) = \left(\lambda_{\backslash\beta}\right)^\top \left(y_{\backslash\beta} - \tilde{y}_{\backslash\beta}\right) - y_\beta + \tilde{y}_\beta$$

eliminates the support function term in ϕ_λ. Defining $u = \phi_\lambda(\tilde{y})$, we multiply both sides of the non-strict inequality in (5.19) by $\exp(-u - \tilde{y}_\beta + y_\beta)$ to get

$$0 \leq \exp\left(\left(\lambda_{\backslash\beta}\right)^\top \left(y_{\backslash\beta} - \tilde{y}_{\backslash\beta}\right)\right) - \exp(-u - \tilde{y}_\beta + y_\beta). \tag{5.20}$$

Convexity of the exponential function tells us that

$$\exp\left(\left(\lambda_{\backslash\beta}\right)^\top \left(y_{\backslash\beta} - \tilde{y}_{\backslash\beta}\right)\right) \leq \left(\lambda_{\backslash\beta}\right)^\top \exp(y_{\backslash\beta} - \tilde{y}_{\backslash\beta}),$$

where the right-hand-side may be rewritten using the Hadamard product

$$\left(\lambda_{\backslash\beta}\right)^\top \exp(y_{\backslash\beta} - \tilde{y}_{\backslash\beta}) = \left(\lambda_{\backslash\beta} \circ \exp(-\tilde{y}_{\backslash\beta})\right)^\top \exp(y_{\backslash\beta}).$$

Applying these observations to (5.20) gives

$$0 \leq \left(\lambda_{\backslash\beta} \circ \exp(-\tilde{y}_{\backslash\beta})\right)^\top \exp(y_{\backslash\beta}) - (\exp(-u - \tilde{y}_\beta)) \exp(y_\beta). \tag{5.21}$$

Inequality (5.21) is essentially what we need to prove the lemma. Defining $z \in \mathbb{R}^{\mathcal{A}}$ by $z_\alpha = \lambda_\alpha \exp(-\tilde{y}_\alpha)$ for $\alpha \neq \beta$ and $z_\beta = -\exp(-u - \tilde{y}_\beta)$, we have that $0 \leq z^\top \exp y$ for all $\exp y \in C_X(\mathcal{A})^\dagger$. As explained at the beginning of this proof, we appeal to continuity to establish $0 \leq z^\top v$ for all $v \in C_X(\mathcal{A})^\dagger$. One may use $\left(\lambda_{\backslash\beta}\right)^\top \mathbf{1} = 1$ to trivially evaluate $z^\top \exp \tilde{y} = 1 - \exp(-u)$, and since $u < 0$ by assumption on $\tilde{y}$, we conclude $z^\top \exp(\tilde{y}) < 0$. $\qquad\square$

Proof of Theorem 5.5.6. By Theorem 5.5.4, we have the dual description $C_X(\mathcal{A})^\dagger = \text{cl}\{\exp y \;:\; (y, 1) \in G_X(\mathcal{A})^\dagger\}$. Applying Theorem 5.5.8 then gives

$$C_X(\mathcal{A})^\dagger = \text{cl}\{\exp y \;:\; \phi_\lambda(y) \geq 0 \;\; \forall \lambda \in \Lambda_X^\star(\mathcal{A})\}.$$

We rewrite the condition on $\phi_\lambda(y)$ as a condition on $v = \exp y$ using the power-cone formulation in Proposition 5.4.7. Since X is polyhedral, Theorem 5.3.7 tells us there are finitely many normalized X-circuits $\Lambda_X(\mathcal{A})$. We may therefore express $C_X(\mathcal{A})^\dagger$ as a finite intersection of dual λ-witnessed AGE cones,

$$C_X(\mathcal{A})^\dagger = \bigcap_{\lambda \in \Lambda_X^\star(\mathcal{A})} C_X(\mathcal{A}, \lambda)^\dagger.$$

Moreover, each dual λ-witnessed AGE cone $C_X(\mathcal{A}, \lambda)^\dagger$ is an outer-approximation of the full-dimensional moment cone $\text{co}\{\exp(\mathcal{A}x) \;:\; x \in X\}$, hence there exists a point v_o in the interior of the moment cone where $v_o \in \text{int} \, C_X(\mathcal{A}, \lambda)^\dagger$ for all $\lambda \in \Lambda_X^\star(\mathcal{A})$. Therefore, by [65, Corollary 16.4.2] we have

$$C_X(\mathcal{A}) = (C_X(\mathcal{A})^\dagger)^\dagger = \sum_{\lambda \in \Lambda_X^\star(\mathcal{A})} (C_X(\mathcal{A}, \lambda)^\dagger)^\dagger = \sum_{\lambda \in \Lambda_X^\star(\mathcal{A})} C_X(\mathcal{A}, \lambda),$$

which establishes the first part of the theorem.

For the second part of the theorem, suppose Λ is a proper subset of $\Lambda_X^\star(\mathcal{A})$. Consider the set $C = \sum_{\lambda \in \Lambda} C_X(\mathcal{A}, \lambda)$ and its dual $C^\dagger = \bigcap\{C_X(\mathcal{A}, \lambda)^\dagger \;:\; \lambda \in \Lambda\}$. Clearly, since $C \subset C_X(\mathcal{A})$ we have $C^\dagger \supset C_X(\mathcal{A})^\dagger$ – we will show that this containment is *strict*, i.e., $C^\dagger \supsetneq C_X(\mathcal{A})^\dagger$. Once this is done, duality will tell us that $C \subsetneq C_X(\mathcal{A})$.

Since C is contained within the signomial nonnegativity cone we again have that $C^\dagger$ contains the moment cone, and so by Lemma 2.6.4 we have $C^\dagger = \text{cl}(C^\dagger \cap \mathbb{R}_{++}^{\mathcal{A}})$. Work with $C^\dagger$ over the positive orthant using Proposition 5.4.7 to express it as $C^\dagger = \text{cl}\{\exp y \;:\; y \in Y\}$ for $Y := \{y \;:\; \phi_\lambda(y) \geq 0 \;\; \forall \lambda \in \Lambda\}$. By Lemma 5.5.11 there exists an element $\tilde{y} \in Y$ for which some $\lambda \in \Lambda_X^\star(\mathcal{A}) \setminus \Lambda$ satisfies $\phi_\lambda(\tilde{y}) < 0$. Apply Lemma 5.5.12 to this pair $(\phi_\lambda, \tilde{y})$ to see that $\exp \tilde{y}$ can be separated from the closed convex set $C_X(\mathcal{A})^\dagger$. We have therefore found a point $\tilde{y}$ where $\exp \tilde{y} \in C^\dagger$ and yet $\exp \tilde{y}$ can be separated from $C_X(\mathcal{A})^\dagger$, so we conclude $C^\dagger \supsetneq C_X(\mathcal{A})^\dagger$. $\qquad \square$

Before concluding this section we would like to point out a more general way to frame our analysis. Given a pair $(\lambda, a) \in \mathbb{R}^m \times \mathbb{R}$ where λ sums to zero and has exactly one negative component $\lambda_i = -1$, we have a power cone constraint

$v_i \leq \exp(a) \prod_{j \neq i} v_j^{\lambda_j}$ which may be rewritten to $1 \leq \exp(a) v^{\lambda}$. Given a finite set of such pairs $\mathsf{P} \subset \mathbb{R}^m \times \mathbb{R}$, we obtain the convex set

$$\mathsf{F}(\mathsf{P}) = \left\{ v \in \mathbb{R}_+^m : 1 \leq \exp(a) v^{\lambda} \; \forall \, (\lambda, a) \in \mathsf{P} \right\}.$$

We have effectively shown that if $\mathsf{K} = \mathrm{co}(\mathsf{P} \cup \{(\mathbf{0}, 1)\})$ is pointed and $\mathsf{F}(\mathsf{P})$ intersects the positive orthant, then the unique minimum $\mathsf{P}^{\star} \subset \mathsf{P}$ for which $\mathsf{F}(\mathsf{P}^{\star}) = \mathsf{F}(\mathsf{P})$ can be read off from the extreme rays of the polyhedral cone K.

5.6 Extreme rays of half-line SAGE cones

In the previous section, we showed that by appropriate appeals to convex duality, one may derive representations of $C_{\mathsf{X}}(\mathcal{A})$ with little to no redundancy. Here we build upon those results to completely characterize the extreme rays of the X-SAGE cone for the univariate case $\mathsf{X} = [0, \infty)$.

Proposition 5.6.1. *For $\alpha_1 < \cdots < \alpha_m$, the extreme rays of $C_{[0,\infty)}(\{\alpha_1, \ldots, \alpha_m\})$ are:*

(1) $\mathbb{R}_+ \cdot \exp(\alpha_1 x)$,

(2) $\mathbb{R}_+ \cdot \{\exp(\alpha_2 x) - \exp(\alpha_1 x)\}$,

(3) $\mathbb{R}_+ \cdot \{c_{i+1} \exp(\alpha_{i+1} x) + c_i \exp(\alpha_i x) + c_{i-1} \exp(\alpha_{i-1} x) : 2 \leq i \leq m - 1\}$ *with*

$$c_{i+1} > 0, \qquad c_{i-1} > 0, \quad and \quad c_i = - \left(\frac{c_{i-1}}{\lambda_{i-1}} \right)^{\lambda_{i-1}} \left(\frac{c_{i+1}}{\lambda_{i+1}} \right)^{\lambda_{i+1}},$$

where

$$\lambda_{i+1} = \frac{\alpha_i - \alpha_{i-1}}{\alpha_{i+1} - \alpha_{i-1}}, \quad \lambda_{i-1} = \frac{\alpha_{i+1} - \alpha_i}{\alpha_{i+1} - \alpha_{i-1}}, \quad and \quad \frac{c_{i-1}}{c_{i+1}} \geq \frac{\lambda_{i-1}}{\lambda_{i+1}}.$$

Proof. Let $\mathcal{A} = \{\alpha_1, \ldots, \alpha_m\}$. By Theorem 5.5.6, all edge generators of $C_{[0,\infty)}(\mathcal{A})$ are either monomials or λ-witnessed AGE functions where λ is a reduced $[0, \infty)$-circuit. By Example 5.5.7, $\Lambda_{[0,\infty)}^{\star}(\mathcal{A}) = \{\delta_2 - \delta_1\} \cup \Lambda_{\mathbb{R}}^{\star}(\mathcal{A})$. Since $n = 1$, Proposition 5.3.4 says all circuits λ have $|\operatorname{supp} \lambda| \leq 3$. We therefore divide the proof into considering cases of monomials, and X-AGE functions with two or three terms.

First we address the monomials. Given $f(x) = \exp(\alpha_i x)$ with $i > 1$, we can write $f = f_1 + f_2$ with $f_1(x) = \exp(\alpha_i x) - \exp(\alpha_{i-1} x)$ and $f_2(x) = \exp(\alpha_{i-1} x)$ – the summand f_1 is nonnegative on $[0, \infty)$ because $\alpha_i > \alpha_{i-1}$, and f_2 is globally

nonnegative. Therefore the only possible extremal monomial in $C_{[0,\infty)}(\mathcal{A})$ is $f(x) = \exp(\alpha_1 x)$. Since $\mathsf{X} = [0, \infty)$, the leading term of any $g \in C_{\mathsf{X}}(\mathcal{A})$ must have a positive coefficient. Moreover, if g is not proportional to f, the leading term of g must have an exponent greater than α_1. Therefore any convex combination of AGE functions $g \in C_{[0,\infty)}(\mathcal{A})$ which are not proportional to f must disagree with $f(x)$ in the limit as x tends to infinity. We conclude f is extremal in $C_{[0,\infty)}(\mathcal{A})$.

Now we consider the 2-term case, where, by Example 5.5.7, we have to consider signomials of the form $f(x) = c_2 \exp(\alpha_2 x) - c_1 \exp(\alpha_1 x)$. We observe that f is nonnegative on $[0, \infty)$ if and only if $c_2 \geq c_1 \geq 0$, and furthermore that such signomials are nonextremal unless $c_1 = c_2$. To see that $f(x) = \exp(\alpha_2 x) - \exp(\alpha_1 x)$ is indeed extremal, note that f cannot be written as a convex combination involving any 3-term AGE functions, because any conic combination of 3-term AGE functions has a leading term with positive coefficient on $\exp(\alpha_i x)$ for some $i \geq 3$.

We have already proven cases (1) and (2) of the proposition. Using Example 5.5.7, we know that any extremal 3-term X-AGE function belongs to a λ-witnessed AGE cone where λ is a reduced $\mathbb{R}$-circuit. These reduced $\mathbb{R}$-circuits have the property $\operatorname{supp} \lambda = \{i-1, i, i+1\}$ $\alpha_{i-1} \lambda_{i-1} + \alpha_{i+1} \lambda_{i+1} = \alpha_i$, $\lambda_i = -1$. Any X-AGE function with such a witness is nonnegative on all of $\mathbb{R}$. Therefore any 3-term X-AGE function f that is *extremal* in $C_{[0,\infty)}(\mathcal{A})$ is also extremal in $C_{\mathbb{R}}(\mathcal{A}) \subset C_{[0,\infty)}(\mathcal{A})$, which (by [130, Prop. 4.4]) implies

$$f(x) = c_{i+1} \exp(\alpha_{i+1} x) - \left(\left[\frac{c_{i+1}}{\lambda_{i+1}} \right]^{\lambda_{i+1}} \left[\frac{c_{i-1}}{\lambda_{i-1}} \right]^{\lambda_{i-1}} \right) \exp(\alpha_i x) + c_{i-1} \exp(\alpha_{i-1} x). \quad (5.22)$$

We have arrived at the final phase of proving part (3) of this proposition. By the equality case in the arithmetic-geometric mean inequality and using

$$\exp(\alpha_i x) = \left(\exp(\alpha_{i+1} x)^{\lambda_{i+1}} \right) \left(\exp(\alpha_{i-1} x)^{\lambda_{i-1}} \right),$$

one finds the unique minimizer $x^\star$ for functions (5.22) satisfies

$$\left[\frac{c_{i+1} \exp(\alpha_{i+1} x^\star)}{\lambda_{i+1}} \right] = \left[\frac{c_{i-1} \exp(\alpha_{i-1} x^\star)}{\lambda_{i-1}} \right] \quad \Leftrightarrow \quad x^\star = \ln\left(\frac{c_{i-1}}{c_{i+1}} \frac{\lambda_{i+1}}{\lambda_{i-1}} \right) / (\alpha_{i+1} - \alpha_{i-1}).$$

If $V_i(\lambda, c) := (c_{i-1} \lambda_{i+1}) / (c_{i+1} \lambda_{i-1})$ satisfies $V_i(\lambda, c) < 1$, then $x^\star < 0$ and by continuity we have $\inf\{f(x) : x \geq 0\} > 0$ – hence the condition $V_i(\lambda, c) \geq 1$ is necessary for extremality. Furthermore, if $V_i(\lambda, c) > 1$, then the unique minimizer of f given by (5.22) occurs at $x^\star > 0$. Such f cannot be decomposed as a convex

combination which involves 1-term or 2-term AGE functions (which have $f(x) > 0$ for $x > 0$), and cannot be written as a convex combination consisting solely of 3-term AGE functions [130, Proposition 4.4], therefore any f given by (5.22) with $V_i(\lambda, c) > 1$ is extremal in $C_{[0,\infty)}(\mathcal{A})$. All that remains is to show extremality of functions (5.22) with $V_i(\lambda, c) = 1$. This follows from the same argument as $V_i(\lambda, c) > 1$, but we must use the stationarity condition $f'(0) = 0$ to preclude using 2-term extremal AGE functions in a decomposition of f.   □

5.7   Discussion and conclusion

In this chapter we have introduced a convex-geometric notion of an X-circuit, which mediates a relationship between point sets $\mathcal{A} \subset \mathbb{R}^n$ and convex sets $\mathsf{X} \subset \mathbb{R}^n$. By showing that this notion of an X-circuit allows an alternative construction of X-SAGE cones (Theorems 5.4.4 and 5.5.5) which cannot be relaxed (Theorem 5.5.6), we have demonstrated that conditional SAGE cones exhibit a substantially richer theory than ordinary SAGE cones. An essential property of this theory is that for general sets X it is not possible to recover an X-circuit $\lambda \in \Lambda_\mathsf{X}(\mathcal{A}, \beta)$ given only information on the signs of its components. As a consequence of this last point – it is not possible to arrive at the concept of conditional SAGE certificates while relying on a "circuit number" approach using only the support of a given polynomial or signomial.

Two lines of theoretical investigations stand out for future work. First, there is the task of formally situating X-circuits in the context of matroid theory (in the case when X is a polyhedron). Here one can use an interpretation from Theorem 5.3.7, that X-circuits $\lambda \in \Lambda_\mathsf{X}(\mathcal{A}, \beta)$ are outer normal vectors to facets of $-\mathcal{A}\mathsf{X} + \mathsf{N}_\beta^\circ$. A broader area of follow-up work is in-depth analysis of multiplicatively-convex sets $\mathsf{S} \subset \mathbb{R}_+^m$ for which $\log(\mathsf{S}) = \{t \ : \ \exp t \in \mathsf{S}\}$ is convex. Some properties of this class of sets include closure under intersection, and closure under the induced-cone operation.

It is of interest to explore the use of the cones $C_\mathsf{X}(\mathcal{A}, \lambda)$ when λ is not an X-circuit. Given a signomial $\sum_{\alpha \in \mathcal{A}} c_\alpha e^\alpha$ with numerical X-SAGE certificate $\{(c^{(\beta)}, v^{(\beta)})\}_{\beta \in \mathcal{A}}$, $c^{(\beta)} \in C_\mathsf{X}(\mathcal{A}, \beta)$, $c \approx \sum_{\beta \in \mathcal{A}} c^{(\beta)}$, one could refine this certificate to higher precision by solving the power-cone program to decompose c as a sum of vectors in $C_\mathsf{X}(\mathcal{A}, \lambda^{(\beta)})$ for $\lambda^{(\beta)} = v^{(\beta)}/|v_\beta^{(\beta)}|$. This would be helpful for large scale problems where $\{(c^{(\beta)}, v^{(\beta)})\}_{\beta \in \mathcal{A}}$ is computed with a first-order solver, or when X is an especially complicated spectrahedron. In the latter case, the standard description of $C_\mathsf{X}(\mathcal{A})$ would be a mixed semidefinite and relative entropy program, while the

formulations for $C_X(\mathcal{A}, \lambda^{(\beta)})$ would be pure power cone programs.

The two obstacles to using Theorem 5.5.6 in computation are that $|\Lambda_X^\star(\mathcal{A})|$ can be exponential in $|\mathcal{A}|$ even when $X = \mathbb{R}^n$, and that finding X-circuits requires a procedure to identify extreme rays of a polyhedral cone. It is not known how severe this first problem is in practice. For the second problem one could focus on X-SAGE polynomials (see Chapter 7) where $X = [-1, 1]^n$ or $X = [0, 1]^n$. The cones of such polynomials supported on $\mathcal{A} \subset \mathbb{N}^n$ are represented by $C_Y(\mathcal{A})$ for $Y = \{y \in \mathbb{R}^n : y \leq 0\}$, and finding $\Lambda_Y^\star(\mathcal{A})$ is made easier by the fact that Y is a cone. The main benefit of this approach for polynomials is the prospect of computing conditional SAGE decompositions in exact arithmetic, especially for sparse polynomials of high degree.

C h a p t e r   6

AN ALGEBRAIC APPROACH TO SIGNOMIAL OPTIMIZATION

6.1   Introduction

Consider the problem of computing

$$f_K^\star = \inf_{x \in K} f(x) \quad \text{where} \quad K = \{x \in X : G(x) \geq 0\} \tag{6.1}$$

for a signomial f, a signomial map G, and a convex set X. We have seen in Chapters 3 and 4 how SAGE certificates provide a mechanism for computing lower bounds on $f_K^\star$. We particularly touched on the idea twice that these bounds are not one-shot procedures; hierarchies of convex relaxations can be devised with the aim of producing arbitrarily strong bounds on $f_K^\star$. The idea of hierarchies of lower bounds is part of a long tradition in the polynomial optimization community and has also been studied for SAGE signomials [13, 124].

A lesser-known segment of the polynomial optimization literature involves using the same nonnegativity certificates to compute *upper bounds* on the minimum of a polynomial optimization problem. Abstractly, this approach can be understood as minimizing the linear function $\mu \mapsto \int f \mathrm{d}\mu$ over some family of probability measures μ supported on K. Lasserre pioneered this latter method by parameterizing probability measures in terms of sums of squares (SOS) polynomials and a reference measure supported on K [143]. To the best of our knowledge, this idea has not been studied for signomials.

We make advances on both of these fronts through a concept of signomial rings. Our main theoretical contributions are a *Positivstellensatz* (a "positive locus theorem") for conditional SAGE and an elementary signomial moment theory. These results lead to hierarchies of REP relaxations that approach the value of a signomial program from both above and below. We focus primarily on the Positivstellensatz, which characterizes signomials positive on sets K given by the intersection of a compact convex set X with preimages of finitely many signomial inequality constraints (as in Equation (6.1)). Through our experiments we have identified the root cause of a known numerical difficulty in certain SAGE relaxations; we show how this difficulty can be overcome by a simple shift of coordinate system.

6.1.1 Summary of results

Throughout this chapter $\mathcal{A} \subset \mathbb{R}^n$ is a distinguished finite ground set that contains the origin. Every vector $\boldsymbol{\alpha} \in \mathcal{A}$ is associated with a "monomial" basis function $e^{\alpha} : \mathbb{R}^n \to \mathbb{R}$ that takes values $e^{\alpha}(\boldsymbol{x}) = \exp\langle \boldsymbol{\alpha}, \boldsymbol{x} \rangle$.

In Section 6.2 we introduce the *signomial ring* $\mathbb{R}[\mathcal{A}]$ as the set of all finite products and real-linear combinations of basis functions $\{e^{\alpha}\}_{\alpha \in \mathcal{A}}$. This concept is simply an organizing framework. It gets us thinking signomials as "polynomial in $\{e^{\alpha}\}_{\alpha \in \mathcal{A}}$," and leaves it to us to account for the fact that formal indeterminates $y_{\alpha} := e^{\alpha}(\boldsymbol{x})$ are not independent of one another. Our techniques show that conditional SAGE has precisely the features necessary to account for the dependence of these monomials, even in the case of irrational exponents.

Our first main result is a characterization of signomials that are positive over sets $K = \{\boldsymbol{x} \in X : g(\boldsymbol{x}) \geq 0 \text{ for all } g \in G\}$ where X is compact and convex and G is a finite set of signomials. To describe the result we need some basic terminology. A signomial is called X-*SAGE* if it can be written as a sum of X-nonnegative signomials each with at most one negative term. A *posynomial* is a signomial with only nonnegative terms. Theorem 6.3.1 states that if $\{f\} \cup G \subset \mathbb{R}[\mathcal{A}]$, then f is positive on K only if there exists an $r \in \mathbb{N}$ for which

$$\left(\sum_{\alpha \in \mathcal{A}} e^{\alpha} \right)^r f = \lambda_f + \sum_{g \in G} \lambda_g \cdot g,$$

where $\lambda_f \in \mathbb{R}[\mathcal{A}]$ is X-SAGE and each $\lambda_g \in \mathbb{R}[\mathcal{A}]$ is a posynomial. Theorem 6.3.1 is the first signomial Positivstellensatz that does not require rational exponents and it is the first conditional SAGE Positivstellensatz that permits nonconvex constraints.

Section 6.4 uses Theorem 6.3.1 as the basis for a hierarchy of REP relaxations to approach the minimum of a signomial f over a set K from below. The statement of the hierarchy has some peculiarities stemming from properties of signomial rings discussed in Subsection 6.2.2. From a performance perspective, these peculiarities work to our benefit by providing stronger bounds at lower levels of the hierarchy. We provide illustrative comparisons of these REP relaxations to SDP relaxations based on the moment-SOS approach and the global solvers BARON [144], ANTIGONE [145], LINDO [146], and SCIP [147, 148].

We turn to signomial moment problems in Sections 6.5 and 6.6. We prove a signomial Riesz-Haviland theorem that, when combined with a moment-determinacy result, leads to a hierarchy of REP relaxations for approaching the value of a signomial minimization problem from above. The hierarchy of upper bounds draws

inspiration from analogous SDP-based hierarchies for polynomial optimization. We find it useful to frame the idea as follows: we fix a reference measure μ with $\operatorname{supp} \mu = \mathsf{K}$, and then minimize the linear function $\phi \mapsto \int f \phi \, d\mu$ over families of signomials ϕ that are nonnegative on K and satisfy $\int \phi \, d\mu = 1$. By allowing sufficiently large classes of such signomials ϕ we can approximate a Dirac distribution centered on a signomial's minimizer on K (although there are of course some technicalities involved).

6.1.2   Related work

We begin by speaking to the broader literature on certifying nonnegativity via the arithmetic-geometric inequality. Beyond SAGE, this includes Reznick's *agiforms* [14], Pantea, Koeppl, and Craciun's characterization of $\mathbb{R}_+^n$-nonnegative circuit polynomials [17], and Iliman and de Wolff's *sums of nonnegative circuits* or *SONC* [96]. The SAGE-SONC relationship was initially unclear, but was resolved in [59, §5] with the introduction of *SAGE polynomials*. One can also understand the SAGE-SONC relationship by implicitly reading results by Wang [100]. Essentially, the unifying perspective is that both of these approaches characterize elementary nonnegative functions that are a sum of "monomials" where at most one monomial contributes a negative value to the overall sum. Further generalizations through this perspective can be found in [130] (which allows monomials $\prod_{i=1}^n |x_i|^{\alpha_i}$) and [60, §4].

Much of the interest in arithmetic-geometric based nonnegativity certificates stems from their sparsity preservation properties [59, 101, 130]. As a consequence of this sparsity preservation, *it is possible to implement these methods with complexity that depends only on the number of terms in the signomial or polynomial* [59]. This property is essential for signomials, since the largest signomial ring $\mathbb{R}[\mathbb{R}^n]$ is not even a unique factorization domain [127]. In the polynomial setting, much work has gone into the development of variants of SOS capable of exploiting *structured sparsity*. See, e.g., [89, 90, 149, 150]. Arithmetic-geometric certificates are significant in the polynomial literature precisely because they can take advantage of *unstructured sparsity*.

Now we turn to some known results for signomial K-nonnegativity problems. To our knowledge, the earliest result here is Delzell's extension of the weak form of Polya's Positivstellensatz to signomials [127]. Chandrasekaran and Shah presented two Positivstellensatz when they introduced SAGE in [13]: one for $\mathsf{K} = \mathbb{R}^n$ and one for Archimedean K. The second of these is based on reduction to Krivine's

Positivstellensatz [111] (see also [112, §5.4.4 (ii)]). Wang, Jaini, Yu, and Poupart developed the first conditional SAGE Positivstellensatz in the case when K is a compact convex set [124]. The signomial exponents must be rational for any of these results to hold. In fact, it is *impossible* to extend Polya's Positivstellensatz to signomials with irrational exponents [127]. We draw comparisons to these results in Subsection 6.3.1.

Our conditional SAGE Positivstellensatz is ultimately driven by reduction to the Dickinson-Povh Positivstellensatz for homogeneous polynomials on $\mathbb{R}^n_+ \setminus \{\mathbf{0}\}$ with infinitely many homogeneous polynomial inequality constraints [151]. Similar appeals were made to this Dickinson-Povh Positivstellensatz in [13] (for $\mathsf{K} = \mathbb{R}^n$) and [124]. Our techniques are distinct from [13, 124] in that our reduction the to Dickinson-Povh Positivstellensatz is with consideration to signomial rings. It is noteworthy that Dickinson and Povh have subsequently used their Positivstellensatz to develop complete hierarchies for *polynomial cone programming* [152]. Our methods can probably be extended to an analogous "signomial cone programming" but we make no attempt to do so here.

Finally we discuss the literature related to our contributions in signomial moment theory. The most concrete connections are in our hierarchy of upper bounds for signomial minimization problems. Indeed, the idea for that hierarchy is lifted almost directly from a work by Lasserre [143] and a subsequent generalization by de Klerk, Lasserre, Laurent, and Sun [153]. The approach in [143] involves semidefinite programming relaxations, so one can say it uses SOS polynomials from a nonnegativity standpoint. The approach in [153] involves the Cassier-Handelman hierarchy [154, 155] and is conceptually closer to our method. Several investigations have been conducted to determine rates of convergence for these upper bounds under various conditions [156–160]. We leave questions of convergence rates with our method to future work.

6.1.3 Notation and other remarks

Throughout this chapter $\mathsf{X} \subset \mathbb{R}^n$ is a closed convex set with support function $z \mapsto \sigma_{\mathsf{X}}(z) = \sup_{x \in \mathsf{X}} \langle z, x \rangle$. We specify a signomial by a vector of coefficients in $c \in \mathbb{R}^{\mathcal{B}}$ (for any finite $\mathcal{B} \subset \mathbb{R}^n$) and write it in a basis expansion $f = \sum_{\beta \in \mathcal{A}} c_\beta e^\beta$.

Our experiments are conducted with MOSEK 9.2 on a Dell XPS 13 9300, with an Intel Core i7-1065G7 processor (4 cores at 1.30GHz) and 16 GB DDR4 RAM (3733 MT/s).

6.2 Signomial rings

As we explained in Subsection 6.1.1, $\mathcal{A} \subset \mathbb{R}^n$ is a finite set that contains the origin and the signomial ring $\mathbb{R}[\mathcal{A}]$ is the $\mathbb{R}$-algebra generated by the basis functions $\{e^\alpha\}_{\alpha \in \mathcal{A}}$. This section explores a way of grading signomial rings by degree. We begin by defining a sequence of sets

$$\mathcal{A}_d = \left\{ \sum_{\alpha \in \mathcal{A}} w_\alpha \alpha \; : \; w \in \mathbb{N}^{\mathcal{A}}, \; \langle \mathbf{1}, w \rangle \leq d \right\} \quad \text{for} \quad d \geq 1.$$

Where we note that $\mathcal{A}_1 = \mathcal{A}$. Next, we formally define the *support* of a signomial $f = \sum_{\alpha \in A} c_\alpha e^\alpha$ as $\text{supp}(f) = \{\alpha \; : \; c_\alpha \neq 0, \; \alpha \in A\}$. The $\mathcal{A}$-*degree* of f is then the smallest integer d for which $\text{supp}(f) \subset \mathcal{A}_d$; this number is denoted $\deg_{\mathcal{A}}(f)$. We use $\mathbb{R}[\mathcal{A}]_{\leq d}$ for the space of signomials of $\mathcal{A}$-degree at most d.

The definition of $\mathcal{A}$-degree is, by itself, enough to get through the proof of our Positivstellensatz in Section 6.3. In later sections it is important to understand certain properties of $\mathcal{A}$-degree. We explore those basic properties here.

6.2.1 The $\mathcal{A}$-degree of a single signomial

The concept of $\mathcal{A}$-degree is artificially imposed on signomials. If $\text{supp}(f) \subset \mathcal{A}$, then the $\mathcal{A}$-degree of f is trivially one. Note that unless $\mathcal{A}$ is decided by some external factor, one can always update $\mathcal{A} \leftarrow \mathcal{A} \cup \text{supp}(f)$, and so every signomial has degree one when considered in a suitable ring. In fact, if we chose to interpret previous Positivstellensatz and hierarchies of SAGE relaxations in terms of signomial rings, then we find that they always make such a choice for $\mathcal{A}$.

In this chapter we show it can be advantageous to consider signomials in rings where their resulting $\mathcal{A}$-degree is greater than one. This creates a need to determine $\deg_{\mathcal{A}}(f)$ when there is no special relationship between $\text{supp}(f)$ and $\mathcal{A}$. The naive thing to do in this case is to explicitly find the smallest ℓ for which $\text{supp}(f) \subset \mathcal{A}_\ell$, but that algorithm does not terminate if f does not belong to $\mathbb{R}[\mathcal{A}]$. In practice we suggest $\mathcal{A}$-degree and membership in $\mathbb{R}[\mathcal{A}]$ be determined by solving an integer-linear program. The natural formulation is given as follows, with $\mathcal{B} = \text{supp}(f)$:

$$\deg_{\mathcal{A}}(f) = \inf\{ \ell \; : \ell \geq 1, \; W \in \mathbb{N}^{\mathcal{A} \times \mathcal{B}} \text{ satisfy}$$

$$\textstyle\sum_{\alpha \in \mathcal{A}} \alpha W_{\alpha\beta} = \beta \text{ and}$$

$$\textstyle\sum_{\alpha \in \mathcal{A}} W_{\alpha\beta} \leq \ell \text{ for all } \beta \text{ in } \mathcal{B}\}.$$

This formulation can be used either directly or in a separable calculation $\deg_{\mathcal{A}}(f) = \max\{\deg_{\mathcal{A}}(e^\alpha) \; : \; \alpha \in \text{supp}(f)\}$. There are many simple ways to improve the

efficiency of computing $\mathcal{A}$-degree, but we do not dwell on this any further. Rather, we consider an example.

Example 6.2.1. Suppose $M \in \mathbb{R}^{n \times n}$ is a dense symmetric matrix and consider the polynomial $p(t) = \prod_{i=1}^{n} t_i + t^\top M t$. From p we construct the signomial f defined by $f(x) = p(\exp x)$. When f is viewed in the rings generated by

$$\mathcal{A} = \{0\} \cup \{\delta_i\}_{i=1}^{n}, \quad \mathcal{A}' = \mathcal{A} \cup \{1_n\}, \quad \text{and} \quad \mathcal{A}'' = \mathcal{A}' \cup \{\delta_i + \delta_j \ : \ (i, j) \in [n]^2\}$$

we have $\deg_{\mathcal{A}}(f) = n$, $\deg_{\mathcal{A}'}(f) = 2$, and $\deg_{\mathcal{A}''}(f) = 1$ respectively.

Although $\mathcal{A}$-degree is not intrinsic to signomials, it exhibits essential properties of coordinate-system invariance. For any b in $\mathbb{R}^n$ and $f \in \mathbb{R}[\mathcal{A}]$, the signomial $g(x) = f(x - b)$ has $\deg_{\mathcal{A}}(g) = \deg_{\mathcal{A}}(f)$. This shift invariance becomes valuable when we discuss numerical optimization in Section 6.4. In addition, for any nonsingular matrix $B \in \mathbb{R}^{n \times n}$, the signomial g defined by $g(x) = f(Bx)$ has $\deg_{[\mathcal{A}B]}(g) = \deg_{\mathcal{A}}(f)$. These invariants are reflected in our proof techniques in Sections 6.3 and 6.5, which are unaffected by changes to $(\mathcal{A}, X)$ that preserve the linear image $\mathcal{A}X \subset \mathbb{R}^{\mathcal{A}}$ up to a translation in the range of $\mathcal{A}$.

6.2.2 Behavior of $\mathcal{A}$-degree under multiplication

Polynomial rings enjoy a property where given two nonzero polynomials p and q, the degree of the product pq is equal to the sum of degrees $\deg p$ and $\deg q$. Signomial $\mathcal{A}$-degree partly preserves this property. For any two signomials f, g in a common ring $\mathbb{R}[\mathcal{A}]$, we have

$$\deg_{\mathcal{A}}(fg) \leq \deg_{\mathcal{A}}(f) + \deg_{\mathcal{A}}(g). \tag{6.2}$$

However, the inequality in (6.2) can be strict even when both f and g are nonzero. A trivial example of strict inequality is given by $f = e^0$, which satisfies $f = f^2$ and $\deg_{\mathcal{A}}(f) = 1$. Here is a nontrivial example.

Example 6.2.2. Consider a positive integer $k \geq 3$ and $\mathcal{A} = \{0, 1, k\}$. Then $f(x) = \exp(x)$ has $\deg_{\mathcal{A}}(f^p) = p$ for $p \in [k - 1]$ and yet $\deg_{\mathcal{A}}(f^k) = 1$.

The potential for strict inequality in (6.2) complicates the process of grading $\mathbb{R}[\mathcal{A}]$ by $\mathcal{A}$-degree. However, this complication can actually be used to our advantage. The idea is that for a polynomial $p \neq 0$, the only polynomial q for which $\deg(pq) < \deg(p)$ is $q = 0$. By contrast, there are certain support sets $\mathcal{A}$, signomials $f \in \mathbb{R}[\mathcal{A}]$, and nontrivial linear subspaces $\mathsf{L} \subset \mathbb{R}[\mathcal{A}]$ where $\deg_{\mathcal{A}}(fg) < \deg_{\mathcal{A}}(f)$ for every signomial $g \in \mathsf{L}$.

Example 6.2.3. Consider $\mathcal{A} = \{-1, 0, 1, 2\}$ and the signomial $f(x) = \exp(3x)$. Clearly $\deg_{\mathcal{A}}(f) = 2$, and yet $\deg_{\mathcal{A}}(gf) \leq 1$ for every signomial g in the one-dimensional linear space $\mathsf{L} = \{c \exp(-x) : c \in \mathbb{R}\} \subset \mathbb{R}[\mathcal{A}]$.

In view of Example 6.2.3, we have a need to take a signomial f and describe the inclusion-maximum $\mathcal{B} \subset \mathcal{A}_d$ where $\deg_{\mathcal{A}}(fg) \leq d$ for every signomial g supported on $\mathcal{B}$. We denote this set by $\operatorname{supp}^{-1}_{\mathcal{A}_d}(f)$ and note that it can be expressed as

$$\operatorname{supp}^{-1}_{\mathcal{A}_d}(f) = \{\boldsymbol{\alpha} \in \mathcal{A}_d : \boldsymbol{\alpha} + \operatorname{supp}(f) \subset \mathcal{A}_d\}. \tag{6.3}$$

In terms of these support sets, Inequality (6.2) simply tells us that if f is of an $\mathcal{A}$-degree k strictly smaller than d, then $\operatorname{supp}^{-1}_{\mathcal{A}_d}(f)$ contains $\mathcal{A}_{d-k}$.

Of course, $\mathcal{A}$-degree can behave like polynomial degree in certain situations. Here is one prominent case.

Proposition 6.2.4. *Suppose f is a nonconstant signomial in $\mathbb{R}[\mathcal{A}]$. If all extreme points of the convex hull of $\mathcal{A}_d$ are among the support of a signomial $g \in \mathbb{R}[\mathcal{A}]_{\leq d}$, then $\deg_{\mathcal{A}}(gf) = d + \deg_{\mathcal{A}}(f)$. In particular, the $\mathcal{A}$-degree of $(\sum_{\boldsymbol{\alpha} \in \mathcal{A}} \mathrm{e}^{\boldsymbol{\alpha}})^d f$ is equal to $d + \deg_{\mathcal{A}}(f)$ whenever f is nonconstant.*

6.3 A Positivstellensatz

Throughout this section f is a signomial in $\mathbb{R}[\mathcal{A}]$, X is a compact convex subset of $\mathbb{R}^n$, and $G \subset \mathbb{R}[\mathcal{A}]$ is finite. Here was present a characterization of signomials that are positive on sets

$$\mathsf{K} = \{\boldsymbol{x} \in \mathsf{X} : g(\boldsymbol{x}) \geq 0 \text{ for all } g \text{ in } G\}.$$

In later sections, this characterization will be used to develop hierarchies of successively stronger convex relaxations for approaching $f_{\mathsf{K}}^{\star} = \inf_{\boldsymbol{x} \in \mathsf{K}} f(\boldsymbol{x})$ from below and above (§6.4 and §6.6 respectively).

Theorem 6.3.1. *If f is positive on K, then there exists an $r \in \mathbb{N}$ for which*

$$\left(\textstyle\sum_{\boldsymbol{\alpha} \in \mathcal{A}} \mathrm{e}^{\boldsymbol{\alpha}}\right)^r f = \lambda_f + \textstyle\sum_{g \in G} \lambda_g \cdot g, \tag{6.4}$$

where $\lambda_f \in \mathbb{R}[\mathcal{A}]$ is X-SAGE and the $\lambda_g \in \mathbb{R}[\mathcal{A}]$ are posynomials.

Note how the theorem requires f to be *positive* on K in order to guarantee an identity that only implies *nonnegativity* on K. The gap between K-positive signomials and K-nonnegative signomials is important in optimization, as it makes the difference

between finite versus asymptotic convergence of our lower bounds. To improve one's chances of finding an identity like (6.4) when $f_K^\star = 0$, the multipliers $(\lambda_g)_{g \in G}$ can be taken as X-SAGE signomials rather than merely posynomials. There is also an interesting case when $G = \emptyset$, where the representation from Theorem 6.3.1 uses no multipliers whatsoever.

Corollary 6.3.2. *If f is positive on* X*, then there exists a natural number r where the signomial $(\sum_{\alpha \in \mathcal{A}} e^\alpha)^r f$ is* X*-SAGE.*

Our proof of Theorem 6.3.1 is presented in Subsection 6.3.2; it relies on two black-box lemmas, which are proven in Subsections 6.3.3 and 6.3.4. The second of these lemmas contains our main technical innovation outside the use of signomial rings, and we provide some extra commentary on the lemma following its proof.

6.3.1 Comparison to existing Positivstellensatz

Here we paraphrase two existing SAGE Positivstellensatz in the language of signomial rings. The first such Positivstellensatz was proven in [13] when the concept of SAGE certificates was introduced. To state the result we use $R_q(G) = \{\prod_{i=1}^q g_i : g_i \in \{1\} \cup G\}$, where "1" refers to the constant signomial e^0.

Theorem 6.3.3 ([13])**.** *Suppose the exponents $\mathcal{A}$ are rational and that each signomial in $\{f\} \cup G$ has $\mathcal{A}$-degree equal to one. Further, assume that G explicitly includes signomials $\{U - e^\alpha, e^\alpha - L\}_{\alpha \in \mathcal{A}}$ for some positive constants U, L, so that* $X = \{x : U \geq e^\alpha(x) \geq L \,\forall\, \alpha \in \mathcal{A}\}$ *is compact. If f is positive on* K*, then there exists a natural number q and a set of $\mathbb{R}^n$-SAGE signomials $(\lambda_h)_{h \in R_q(G)} \subset \mathbb{R}[\mathcal{A}]$ that satisfy $f = \sum_{h \in R_q(G)} \lambda_h \cdot h$.*

We have phrased Theorem 6.3.3 to make clear that if (f, G) satisfy its hypothesis then they also satisfy the hypothesis of Theorem 6.3.1 for the indicated choice of X. Theorem 6.3.1 is qualitatively different from Theorem 6.3.3 in that the former does not require taking products of constraint functions; this distinction is of practical importance when working with constraint signomials of $\mathcal{A}$-degree greater than one.

The next Positivstellensatz was proven by Wang et al. [124] shortly after the introduction of conditional SAGE certificates. Its scope is limited to problems where $G = \emptyset$ (i.e., K = X), but is nevertheless distinguished in how its conclusion is independent of the representation of X.

Theorem 6.3.4 ([124])**.** *Suppose the exponents $\mathcal{A}$ are rational and that f has $\mathcal{A}$-degree one. If f is positive on* X, *then there exists a natural number r for which $(\sum_{\alpha \in \mathcal{A}} e^{\alpha})^r f$ is* X-*SAGE.*

Our Theorem 6.3.1 naturally generalizes Wang at al.'s Theorem 6.3.4 to the constrained setting, and in fact our proof of Theorem 6.3.1 draws much inspiration from [124]. The comparison between Wang et al.'s Theorem 6.3.4 and our Corollary 6.3.2 is best illustrated with an example.

Example 6.3.5. Return to the signomials from Example 6.2.1. Let M_{ij} denote the entries of the matrix $\boldsymbol{M}$ so the signomial $f = e^{\boldsymbol{1}} + \sum_{i,j=1}^{n} M_{ij} e^{[\delta_i + \delta_j]}$ is positive on X. If we want a certificate that f is nonnegative over X, then Theorem 6.3.4 says it suffices to look for X-SAGE decompositions of functions $\mathcal{L}_r = \left(e^{\boldsymbol{0}} + e^{\boldsymbol{1}} + \sum_{i \leq j}^{n} e^{[\delta_i + \delta_j]} \right)^r f$. Since the number of terms in $\mathcal{L}_r$ grows as $O(n^{2r})$, the sizes of the REPs used when searching for the X-SAGE decompositions can scale as rapidly as $O(n^{4r})$. By contrast, Corollary 6.3.2 says it suffices to look for X-SAGE decompositions of functions $\mathcal{L}'_r = \left(\sum_{\alpha \in \mathcal{A}} e^{\alpha} \right)^r f$ where $\mathcal{A}$ is *any* set for which f belongs to $\mathbb{R}[\mathcal{A}]$. In particular, we can use $\mathcal{A} = \{\boldsymbol{0}\} \cup \{\boldsymbol{\delta}_i\}_{i=1}^n$, so the number of terms in $\mathcal{L}'_r$ would grow as only $O(n^r) \ll O(n^{2r})$. Corollary 6.3.2 therefore justifies a whole family of convergent convex relaxation hierarchies with different efficiency profiles as the hierarchy parameter increases.

Besides the comparisons we have drawn so far, we make no requirement that the exponents $\mathcal{A}$ are rational. The distinction between rational and irrational exponents has some mathematical significance. In 2008, Delzell studied the extent to which Polya's theorem (for homogeneous polynomials positive on the simplex) generalizes to signomials in $\mathbb{R}[\mathbb{R}^n]$ [127]. Using the convention of signomials as functions $\boldsymbol{t} \mapsto \sum_{\alpha} c_{\alpha} \boldsymbol{t}^{\alpha}$, [127] showed that the bivariate signomial $f(\boldsymbol{t}) = t_1^2 + t_2^2 - t_1^{1+\epsilon} t_2^{1-\epsilon}$ is positive on $\mathbb{R}_{++}^2$ when $\epsilon \in (-1, 1)$, and yet when ϵ is irrational, there exists no "homogeneous" signomial $g \in \mathbb{R}[\mathbb{R}^n]$ for which gf has nonnegative coefficients. That is, it is impossible to generalize even the weak form of Polya's theorem to signomials with irrational exponents. Our results show that under a different model of signomial rings and a compactness assumption, X-SAGE certificates characterize signomials positive on X even when the exponents are irrational.

6.3.2 Proof of Theorem 6.3.1

Our proof works by mapping a signomial problem to a polynomial problem, applying a polynomial Positivstellensatz, and then mapping back to signomials. As a first

step long this path, we shall call a polynomial p a *polynomialization* of f if $f(x) = p(\exp \mathcal{A}x)$ for all x in $\mathbb{R}^n$. Note that every signomial f has a homogeneous polynomialization of degree $\deg_{\mathcal{A}}(f)$ (since $\mathbf{0} \in \mathcal{A}$). Henceforth, we assume all polynomializations are homogeneous.

Example 6.3.6. Consider the univariate case $\mathcal{A} = [1/4; 1/2; 1/3; 0] \in \mathbb{R}^{4\times 1}$. The signomial $f(x) = \exp(x)$ admits several polynomializations, among them $p_1(y) = y_1^2 y_2$ and $p_2(y) = y_3^3$. Note that $p_1 \cong p_2$ on the variety $\{y \in \mathbb{R}^4 : y_1^2 = y_2, y_2^2 = y_3^3\}$.

The above example suggests a signomial ring $\mathbb{R}[\mathcal{A}]$ is equivalent to the ring of polynomials on $\mathbb{R}^{\mathcal{A}}$, modulo a suitable binomial ideal to capture the relationships between $\alpha, \beta \in \mathcal{A}$. In order to interpret a signomial ring in this way, we need $\exp(\mathcal{A}\mathbb{R}^n)$ to be the intersection of a toric variety with a positive orthant. By considering $\mathcal{A} = \{1, \sqrt{2}, 0\}$ we see that this cannot be the case in general. This provides one example of how fully general signomial rings are resistant to techniques from traditional algebraic geometry. However, the differences between signomial and polynomial rings are less pronounced when considering these functions only over compact sets. Specifically, by restricting our attention to signomial nonnegativity on compact sets, we are able to prove Theorem 6.3.1 by appeals to the following results of Dickinson and Povh.

Theorem 6.3.7 ([151])**.** *Let p be a homogeneous polynomial on $\mathbb{R}^{\mathcal{A}}$, and let Q be a finite set of homogeneous polynomials on $\mathbb{R}^{\mathcal{A}}$ that includes the constant polynomial $y \mapsto 1$. If p is positive on $\{y \in \mathbb{R}_+^{\mathcal{A}} : q(y) \geq 0 \text{ for all } q \in Q\} \setminus \{\mathbf{0}\}$, then for some $r \in \mathbb{N}$ there exist homogeneous polynomials $\{\mu_q\}_{q \in Q}$ with nonnegative coefficients such that $(\sum_{\alpha \in \mathcal{A}} y_\alpha)^r p(y) = \sum_{q \in Q} \mu_q(y) q(y)$.*

For general choices of $(\mathcal{A}, \mathsf{X})$ we also require a reduction from a semi-infinite nonnegativity problem to a finite nonnegativity problem, as follows.

Theorem 6.3.8 ([151])**.** *Consider a countable set $\{p\} \cup Q$ of homogeneous polynomials on $\mathbb{R}^{\mathcal{A}}$. If p is positive on $\{y \in \mathbb{R}_+^{\mathcal{A}} : q(y) \geq 0 \text{ for all } q \in Q\} \setminus \{\mathbf{0}\}$, then there exists a finite $Q' \subset Q$ for which p is positive on $\{y \in \mathbb{R}_+^{\mathcal{A}} : q(y) \geq 0 \,\forall\, q \in Q'\} \setminus \{\mathbf{0}\}$.*

Next, given a polynomial p on $\mathbb{R}^{\mathcal{A}}$, we have the *signomialization $x \mapsto p(\exp \mathcal{A}x)$*. Signomialization transparently preserves important algebraic properties. For example, if we signomialize a polynomial that has nonnegative coefficients in the monomial basis, then we obtain a posynomial. In addition, if g is the signomialization of a polynomial p and p is a polynomialization of some signomial f, then $g = f$.

The following lemma roughly shows how these concepts help map Dickinson-Povh certificates to conditional SAGE certificates.

Lemma 6.3.9. *Let $Q = Q_1 \cup Q_2$ be a finite set of polynomials on $\mathbb{R}^{\mathcal{A}}$ where the signomialization of each $q \in Q_2$ is X-AGE and suppose p is a polynomialization of f. If $(\sum_{\alpha \in \mathcal{A}} y_\alpha)^r p(\mathbf{y}) = \sum_{q \in Q} \mu_q(\mathbf{y}) q(\mathbf{y})$ for polynomials μ_q with nonnegative coefficients and a natural number r, then there exists an X-SAGE function $\lambda_f \in \mathbb{R}[\mathcal{A}]$ for which $(\sum_{\alpha \in \mathcal{A}} e^\alpha)^r f = \lambda_f + \sum_{q \in Q_1} \lambda_q g_q$, where λ_q is the signomialization of μ_q and g_q is the signomialization of q.*

The work in our proof of Theorem 6.3.1 is to derive polynomial data from signomial data so that the hypotheses of Lemma 6.3.9 are satisfied. Much of this work is accomplished in our next lemma.

Lemma 6.3.10. *There exists a countable set of homogeneous polynomials $Q(\mathsf{X})$ on $\mathbb{R}^{\mathcal{A}}$ satisfying the following properties:*

   (i) each $q \in Q(\mathsf{X})$ has at most two terms,

   (ii) $\exp \mathcal{A}\mathsf{X} = \{\mathbf{y} \in \mathbb{R}^{\mathcal{A}} : q(\mathbf{y}) \geq 0 \ \text{for all} \ q \in Q(\mathsf{X}), \ y_{\mathbf{0}} = 1\}$,

   (iii) if $\mathbf{y}$ is a nonzero vector where $q(\mathbf{y}) \geq 0$ for all $q \in Q(\mathsf{X})$, then $\mathbf{y} > \mathbf{0}$.

As a consequence of conditions (i) and (ii) in the lemma, the signomialization of any $q \in Q(\mathsf{X})$ has at most two terms and is X-nonnegative.

Proof of Theorem 6.3.1. Fix $f > 0$ on $\mathsf{K} := \{\mathbf{x} \in \mathsf{X} : g(\mathbf{x}) \geq 0 \forall g \in G\}$. Let p be a polynomialization of f, $Q(G)$ be a set of polynomializations of G (one polynomial for each signomial in G), and $Q(\mathsf{X})$ be as in Lemma 6.3.10. Define the region $\mathsf{K}_p = \exp \mathcal{A}\mathsf{K}$ within $\mathbb{R}^{\mathcal{A}}$ and the set of polynomials $Q = Q(\mathsf{X}) \cup Q(G)$.

We begin by noting the identity $\mathsf{K}_p = (\exp \mathcal{A}\mathsf{X}) \cap \{\exp \mathcal{A}\mathbf{x} : g(\mathbf{x}) \geq 0 \forall g \in G\}$. Next, we apply Lemma 6.3.10 and we use the fact that $g(\mathbf{x}) = q(\exp \mathcal{A}\mathbf{x})$ when q is a polynomialization of g. This allows us to write K_p purely in terms of homogeneous polynomials: $\mathsf{K}_p = \{\mathbf{y} : y_{\mathbf{0}} = 1, \ q(\mathbf{y}) \geq 0 \forall q \in Q\}$. From here we drop the constraint $y_{\mathbf{0}} = 1$ to obtain $\mathsf{T} = \{\mathbf{y} : q(\mathbf{y}) \geq 0 \forall q \in Q\}$. Apply the third property of $Q(\mathsf{X})$ from Lemma 6.3.10 to see that $\mathsf{T} \setminus \{\mathbf{0}\}$ is contained within $\mathbb{R}^{\mathcal{A}}_{++}$.

Let $d = \deg_{\mathcal{A}}(f)$ and consider an arbitrary vector $\mathbf{y} \in \mathsf{T} \setminus \{\mathbf{0}\}$. Since p is homogeneous, we have $p(\mathbf{y}) = y_{\mathbf{0}}^d p(\mathbf{y}/y_{\mathbf{0}})$. Similarly, because all polynomials defining

T are homogeneous, we have that $\tilde{y} := y/y_0$ is in K_p. By the definition of K_p we know that every vector $\tilde{y} \in \mathsf{K}_p$ can be represented as $\tilde{y} = \exp \mathcal{A}x$ for suitable $x \in \mathsf{K}$. Since p is a polynomialization of f, we find $p(y) = y_0^d f(x)$, which is positive by assumption on f, K. We therefore have that p is positive on $\mathsf{T} \setminus \{\mathbf{0}\}$.

By Theorem 6.3.8 there exists a $Q' \subset Q$ that is finite and where p is positive on $\mathsf{T}' \setminus \{\mathbf{0}\}$ for $\mathsf{T}' := \{y \; : \; q(y) \geq 0 \; \forall \; q \in Q'\}$. We are free to assume $Q' = Q(\mathsf{X})' \cup Q(G)$ where $Q(\mathsf{X})' \subset Q(\mathsf{X})$ includes the constant polynomial $y \mapsto 1$. By Theorem 6.3.7, there exists an $r \in \mathbb{N}$ and homogeneous polynomials $\{h_q\}_{q \in Q'}$ on $\mathbb{R}^{\mathcal{A}}$ with nonnegative coefficients where

$$(\textstyle\sum_{\alpha \in \mathcal{A}} y_\alpha)^r p(y) = \textstyle\sum_{q \in Q'} h_q(y) q(y).$$

From property (i) of $Q(\mathsf{X})$ we know that each constraint polynomial $q \in Q(\mathsf{X})'$ has at most two terms. In addition, property (ii) of $Q(\mathsf{X})$ tells us that the signomialization of any $q \in Q(\mathsf{X})'$ is X-nonnegative. It is easily verified that all X-nonnegative signomials with at most two terms are X-AGE. We may therefore apply Lemma 6.3.9 to obtain

$$(\textstyle\sum_{\alpha \in \mathcal{A}} e^\alpha)^r f = \lambda_f + \textstyle\sum_{q \in Q(G)} \lambda_q g_q$$

for signomials $g_q(x) = q(\exp \mathcal{A}x)$, posynomials $\lambda_q(x) = \mu_q(\exp \mathcal{A}x)$, and an X-SAGE $\lambda_f \in \mathbb{R}[\mathcal{A}]$. We complete the proof by noting that $\{g_q\}_{q \in Q(G)} = G$. $\quad\square$

We emphasize that the decomposition promised in Theorem 6.3.1 makes no reference to the set $Q(\mathsf{X})$ used in our proof of the theorem. This reflects how such a decomposition exists for given $r \in \mathbb{N}$ if (but not only if) there are *any* polynomials $Q(\mathsf{X})$ satisfying Lemma 6.3.9 where the polynomialization of f admits a Dickinson-Povh certificate over $\{y \; : \; q(y) \geq 0 \text{ for all } q \in Q(G) \cup Q(\mathsf{X})\}$ with exponent r. So by virtue of using SAGE certificates we do not need to construct $Q(\mathsf{X})$ explicitly, and in fact we automatically do at least as well as choosing the *best possible* $Q(\mathsf{X})$ consistent with Lemmas 6.3.10 and Theorems 6.3.7 and 6.3.8.

6.3.3 Proof of Lemma 6.3.9

Let $r \in \mathbb{N}$ be such that the stated polynomials μ_q exist, and let λ_q, g_q be the signomializations given in the lemma statement. Since μ_q are polynomials with nonnegative coefficients, the signomializations λ_q are posynomials. Set $\lambda_f = \sum_{q \in Q_2} \lambda_q g_q$. We are given that the signomialization g_q of any $q \in Q_2$ is X-AGE. Since the product of an X-AGE function with a posynomial is X-SAGE, and sums of such products are

likewise X-SAGE, we find that the stated λ_f is X-SAGE. Completing the proof is a matter of purely algebraic identifications. Namely,

$$
\begin{aligned}
(\textstyle\sum_{\alpha \in \mathcal{A}} \mathrm{e}^\alpha(\boldsymbol{x}))^r f(\boldsymbol{x}) &= (\textstyle\sum_{\alpha \in \mathcal{A}} \mathrm{e}^\alpha(\boldsymbol{x}))^r p(\exp \mathcal{A}\boldsymbol{x}) \\
&= \sum_{q \in Q} \mu_q(\exp \mathcal{A}\boldsymbol{x}) q(\exp \mathcal{A}\boldsymbol{x}) = \lambda_f(\boldsymbol{x}) + \sum_{q \in Q_1} \lambda_g(\boldsymbol{x}) g_q(\boldsymbol{x}),
\end{aligned}
$$

where the last equality decomposed the sum over $Q = Q_1 \cup Q_2$ and applied the definitions of λ_f, λ_q, and g_q.

6.3.4 Proof of Lemma 6.3.10

We build up this set of polynomials incrementally. To avoid clutter we use the symbol Q rather than $Q(\mathsf{X})$ for the proof, and we set $\mathsf{T} := \{\boldsymbol{y} : q(\boldsymbol{y}) \geq 0 \,\forall\, q \in Q\}$ for the current value of Q. Begin by initializing $Q = \{\boldsymbol{y} \mapsto y_\alpha : \alpha \in \mathcal{A}\}$, so that $\mathsf{T} = \mathbb{R}_+^{\mathcal{A}}$. Then update Q to include the $2(|\mathcal{A}| - 1)$ linear functions

$$
q_\alpha^+(\boldsymbol{y}) = y_0 \left(\max_{\boldsymbol{x} \in \mathsf{X}} \mathrm{e}^\alpha(\boldsymbol{x}) \right) - y_\alpha \quad \text{and} \quad q_\alpha^-(\boldsymbol{y}) = y_\alpha - y_0 \left(\min_{\boldsymbol{x} \in \mathsf{X}} \mathrm{e}^\alpha(\boldsymbol{x}) \right)
$$

for $\alpha \in \mathcal{A} \setminus \{\boldsymbol{0}\}$. The functions q_α^+ ensure that any vector $\boldsymbol{y} \in \mathsf{T}$ with $y_0 = 0$ necessarily satisfies $\boldsymbol{y} = \boldsymbol{0}$. Conversely, the functions q_α^- ensure that when $y_0 > 0$ we have $\boldsymbol{y} > \boldsymbol{0}$. When considered together, we have that if $\boldsymbol{y}$ is a nonzero vector in T, then $\boldsymbol{y} > \boldsymbol{0}$, and so the set of polynomials Q already satisfies property (iii) in the lemma statement.

We turn to property (ii). Recall the notation where δ_α is the standard basis vector in $\mathbb{R}^{\mathcal{A}}$ corresponding to $\alpha \in \mathcal{A}$. Then $\boldsymbol{y}$ belongs to $\exp \mathcal{A}\mathsf{X}$ if an only if there exists a $\boldsymbol{z} \in \mathcal{A}\mathsf{X}$ where $y_\alpha = \mathrm{e}^{\delta_\alpha}(\boldsymbol{z})$. The set $\mathcal{A}\mathsf{X}$ is compact and convex, therefore by a continuity argument (or a direct application of [161, Theorem 3.1]) there exists a set $\mathsf{S} \subset \mathbb{Z}^{\mathcal{A}} \times \mathbb{R}$ where $\boldsymbol{z} \in \mathcal{A}\mathsf{X}$ holds if and only if $\langle \boldsymbol{a}, \boldsymbol{z} \rangle \leq \log b$ for all $(\boldsymbol{a}, \log b) \in \mathsf{S}$. We can take the set S to be countable by always choosing $\log b = \sigma_{\mathsf{X}}(\mathcal{A}^\dagger \boldsymbol{a})$. For given $(\boldsymbol{a}, \log b) \in \mathsf{S}$, take componentwise maximums $\gamma = \boldsymbol{0} \wedge \boldsymbol{a}$ and $\kappa = \boldsymbol{0} \wedge (-\boldsymbol{a})$, so the inequality $\langle \boldsymbol{a}, \boldsymbol{z} \rangle \leq \log b$ is equivalent to $\mathrm{e}^\gamma(\boldsymbol{z}) \leq b \mathrm{e}^\kappa(\boldsymbol{z})$. For each such signomial inequality there is a polynomial inequality $\boldsymbol{y}^\gamma \leq b \boldsymbol{y}^\kappa$ that is equivalent in the relevant regime $\boldsymbol{y} \in \mathsf{T}$, $y_0 = 1$. Setting $v = \|\boldsymbol{a}\|_1 - \|\kappa\|_1$ and $u = \|\boldsymbol{a}\|_1 - \|\gamma\|_1$, we homogenize the polynomial inequality defined above to an equivalent form $q(\boldsymbol{y}) = b y_0^v \boldsymbol{y}^\kappa - y_0^u \boldsymbol{y}^\gamma \geq 0$. We finalize Q by updating it to contain all homogeneous polynomials obtained in this way. Since all reformulations employed here were reversible over $\mathbb{R}_{++}^{\mathcal{A}}$, we have (ii): $\exp \mathcal{A}\mathsf{X} = \{\boldsymbol{y} : q(\boldsymbol{y}) \geq 0 \,\forall\, q \in Q, y_0 = 1\}$.

As property (i) holds by construction, the proof is complete.

Remark 6.3.11. The exponents of the polynomials in Q were derived from halfspaces that contain the compact convex set $\mathcal{A}X$. Since $\mathcal{A}X$ is low-dimensional in general, some of these halfspaces can come together to form *hyperplanes* containing $\mathcal{A}X$. Let $H \subset \mathbb{N}^{\mathcal{A}}$ denote the set of all integral normal vectors of hyperplanes that contain $\mathcal{A}X$. For $\boldsymbol{a} \in H$ we can use the construction described above to obtain a polynomial $\boldsymbol{y} \mapsto q(\boldsymbol{y}) = b y_0^{\nu} \boldsymbol{y}^{\kappa} - y_0^{u} \boldsymbol{y}^{\gamma}$ where $q(\exp \mathcal{A} \boldsymbol{x}) = 0$ for all $\boldsymbol{x}$ in X. If X is full-dimensional then the real locus of these polynomials is more or less the smallest toric variety that contains $Y := \exp \mathcal{A} \mathbb{R}^n$. It is possible that Y is poorly approximated by a variety; in this case we are leveraging compactness to provide a *local* description for $\exp \mathcal{A}X$ in terms of infinitely many polynomial inequalities. Prior assumptions from [13, 124] that $\mathcal{A} \subset \mathbb{Q}^n$ were used to construct polynomial equations for describing $\exp \mathcal{A} \mathbb{R}^n$ as the intersection of a variety with the positive orthant.

6.4 A complete hierarchy of lower bounds

This section demonstrates how the concept of signomial rings leads to improved methods for lower-bounding and solving nonconvex optimization problems. Formally, given a finite set of signomials $\{f\} \cup G$ and a convex set X, we would like to solve

$$f_K^\star = \inf_{\boldsymbol{x} \in K} f(\boldsymbol{x}) \quad \text{where} \quad K = \{\boldsymbol{x} \in X : g(\boldsymbol{x}) \geq 0 \text{ for all } g \text{ in } G\}. \tag{6.5}$$

Our high-level approach here is quite standard. We want certificates that shifted signomials $f - \gamma$ are nonnegative on K, and such certificates are available to us through Theorem 6.3.1. In order to implement this idea we just need to grade the certificates according to largest $\mathcal{A}$-degree of the constituent signomials.

We recall two essential definitions from Section 6.2. First, the $\mathcal{A}$-degree of a signomial h is the smallest integer ℓ for which $\text{supp}(h) \subset \mathcal{A}_\ell$. Second, for a given signomial h and positive integer d, the set $\mathcal{B} := \text{supp}_{\mathcal{A}_d}^{-1}(h)$ is the largest $\mathcal{B} \subset \mathcal{A}_d$ for which $\deg_{\mathcal{A}}(h e^{\boldsymbol{\beta}}) \leq d$ for every $\boldsymbol{\beta}\mathcal{B}$.

Definition 6.4.1. *Given an integer d where $r := d - \deg_{\mathcal{A}}(f) \geq 0$, the $\mathcal{A}$-degree d SAGE bound for Problem* (6.5) *is*

$$f_K^{(d)} := \sup \ \gamma \ s.t. \ \left(\textstyle\sum_{\alpha \in \mathcal{A}} e^{\alpha} \right)^r (f - \gamma) - \textstyle\sum_{g \in G} \lambda_g g \in C_X(\mathcal{A}_d), \tag{6.6}$$

$$\gamma \in \mathbb{R}, \ and \ \lambda_g \in C_X\left(\text{supp}_{\mathcal{A}_d}^{-1}(g)\right) \ for \ each \ g \in G.$$

When $d < \deg_{\mathcal{A}}(f)$, we set $f_K^{(d)} = -\infty$.

We make no assumptions about the $\mathcal{A}$-degree of constraint signomials; degenerate cases are covered by the fact that $\mathrm{supp}_{\mathcal{A}_d}^{-1}(g)$ can be empty, which in turn forces $\lambda_g = 0$. Note it is possible that $\deg_{\mathcal{A}}(g) > d$ and yet $\mathrm{supp}_{\mathcal{A}_d}^{-1}(g)$ is nonempty. When the hierarchy is applied to problems with an equality constraints $g(\boldsymbol{x}) = 0$, one simply uses an unconstrained multiplier $\lambda_g \in \mathrm{Span}\{e^{\alpha} : \alpha \in \mathrm{supp}_{\mathcal{A}_d}^{-1}(g)\}$.

Corollary 6.4.2. *The sequence* $f_{\mathsf{K}}^{(1)}, f_{\mathsf{K}}^{(2)}, \dots$ *is nondecreasing and bounded above by* $f_{\mathsf{K}}^{\star}$. *If the signomials* $\{f\} \cup G$ *belong to* $\mathbb{R}[\mathcal{A}]$ *and* X *is compact, then*

$$\lim_{d \to \infty} f_{\mathsf{K}}^{(d)} = f_{\mathsf{K}}^{\star}.$$

Proof. The sequence is nondecreasing because $\mathrm{supp}_{\mathcal{A}_d}^{-1}(g) \subset \mathrm{supp}_{\mathcal{A}_{d+1}}^{-1}(g)$ and $C_{\mathsf{X}}(\mathcal{B}) \subset C_{\mathsf{X}}(\mathcal{B}')$ whenever $\mathcal{B} \subset \mathcal{B}' \subset \mathbb{R}^n$. That is, the feasible sets grow with d. The sequence is bounded above by $f_{\mathsf{K}}^{\star}$ because every feasible solution certifies $f(\boldsymbol{x}) \geq \gamma$ for all $\boldsymbol{x} \in \mathsf{K}$. Under the assumptions on X and $\mathbb{R}[\mathcal{A}]$, convergence to $f_{\mathsf{K}}^{\star}$ follows from Theorem 6.3.1 and the fact that posynomials are trivially X-SAGE. $\qquad\square$

Corollary 6.4.2 is the first completeness result for minimizing an arbitrary signomial subject to constraints given by a compact convex set and a conjunction of arbitrary (but finitely many) signomial inequalities. It is also the first completeness result for a hierarchy that uses conditional SAGE certificates in the presence of nonconvex constraints. The approach is also notable because the hierarchy is indexed by a single parameter d (much like a Lasserre-relaxation). By contrast, other SAGE-based hierarchies have been indexed by two or even three parameters. This difference stems from how we decide supports of the generalized Lagrange multipliers λ_g with consideration to the signomial ring $\mathbb{R}[\mathcal{A}]$ and the $\mathcal{A}$-degree of the constraint functions g.

The remainder of this section explores the practicality of our hierarchy through three examples, stated in terms of variables $\boldsymbol{t} = \exp \boldsymbol{x}$. Through the latter two problems it becomes evident that a shift of coordinate system can be crucial in accurately computing $f_{\mathsf{K}}^{(d)}$. We get ahead of these examples in Subsection 6.4.2, where we discuss problem scaling for SAGE relaxations and signomial and polynomial optimization more generally.

6.4.1 Polynomial optimization on the positive orthant

If all signomials in (6.5) have integer exponents and finite lower bounds on the decision variable $\boldsymbol{x}$, then the problem can be written with polynomials in $\boldsymbol{t}$ by

clearing denominators. The following nonconvex quadratic program was obtained by applying this procedure to [162, Problem 23].

$$\min_{t \in \mathbb{R}^5_{++}} \quad 5.3578\,t_3^2 + 0.8357\,t_1 t_5 + 37.2393 t_1 \tag{6.7}$$

$$\text{s.t} \quad \left. \begin{aligned} 0.06663\,t_2 t_5 - 0.02584\,t_3 t_5 + 0.0734\,t_1 t_4 + 1000 &\geq 0 \\ 0.33085\,t_3 t_5 - 0.853007\,t_2 t_5 - 0.09395\,t_1 t_4 + 1000 &\geq 0 \\ 0.4200\,t_1 t_2 + 0.30586\,t_3^2 + t_2 t_5 - 1330.3294 &\geq 0 \\ 0.2668\,t_1 t_3 + 0.40584\,t_3 t_4 + t_3 t_5 - 2275.1326 &\geq 0 \end{aligned} \right\} G_{\text{nonconvex}}$$

$$1000 - 0.24186\,t_2 t_5 - 0.10159\,t_1 t_2 - 0.07379\,t_3^2 \geq 0$$

$$1000 - 0.29955\,t_3 t_5 - 0.07992\,t_1 t_3 - 0.12157\,t_3 t_4 \geq 0$$

$$\left. \begin{aligned} (102, 45, 45, 45, 45) - t &\geq \mathbf{0} \\ t - (78, 33, 27, 27, 27) &\geq \mathbf{0} \end{aligned} \right\} G_{\text{box}}$$

We have labeled the set of inequality constraints that are nonconvex in x as $G_{\text{nonconvex}}$ and use G_{box} for the signomials that imply box constraints on x. We use G_{all} to refer to all constraints appearing in (6.7). By applying solution recovery to the SAGE relaxations discussed below, one can certify that the optimal solution to this problem is $t^\star \approx (78, 33, 29.99574, 45, 36.77533)$ with optimal objective $f_K^\star \approx 10122.4932$.

This problem lets us illustrate the effect of considering different signomial rings and different sets of "algebraic" constraints G. While exploring these effects, we fix

$$\mathsf{X} = \{x \ : \ g(x) \geq 0 \text{ for all } g \in G_{\text{all}} \setminus G_{\text{nonconvex}}\}.$$

We examine three cases where set G to G_{all}, to $G_{\text{all}} \setminus G_{\text{box}}$, and to $G_{\text{nonconvex}}$. For each choice of G we consider two types of signomial rings. For the *naive rings* we take $\mathcal{A}$ as the smallest set so every signomial in $\{f\} \cup G$ has $\mathcal{A}$-degree one. The naive rings have generating sets of size 19, 15, and 12 (as G gets smaller). We also use a *natural ring* $\mathcal{A} = \{\mathbf{0}, \delta_1, \dots, \delta_n\}$ that reflects how (6.7) is polynomial in t.

Performance data for the SAGE relaxations is given in Tables 6.1 and 6.2. We see finite convergence for the hierarchy in four out of the six choices of $(G, \mathcal{A})$. The best bound at each hierarchy level used G_{all}. This reflects a known phenomenon where incorporating a constraint in an explicit algebraic way can improve bounds even when the constraint is nominally accounted for in the set X. The fact that the

presence of G_{box} in G affects finite convergence may be due to how box constraints are binding at optimality for (t_1, t_2, t_4). Another key point from the data is that the solver runtimes scale more gracefully when using the natural ring compared to the naive ring. This is to be expected, since the natural ring is smaller than the naive rings.

Table 6.1: Natural-ring SAGE bounds and solver runtimes for problem (6.7).

d	$\mathcal{A}$-degree d SAGE bounds			solver runtimes (s)		
	G_{all}	$G_{\text{all}} \setminus G_{\text{box}}$	$G_{\text{nonconvex}}$	G_{all}	$G_{\text{all}} \setminus G_{\text{box}}$	$G_{\text{nonconvex}}$
2	10022.940	9322.848	9322.849	0.070	0.015	0.015
3	10122.493	9964.326	9954.832	0.588	0.184	0.132
4	-	10122.493	10074.250	-	1.314	1.368

Table 6.2: Naive-ring SAGE bounds and solver runtimes for problem (6.7).

d	$\mathcal{A}$-degree d SAGE bounds			solver runtimes (s)		
	G_{all}	$G_{\text{all}} \setminus G_{\text{box}}$	$G_{\text{nonconvex}}$	G_{all}	$G_{\text{all}} \setminus G_{\text{box}}$	$G_{\text{nonconvex}}$
1	10022.929	9322.848	9322.849	0.045	0.015	0.015
2	10122.493	10069.946	10059.838	1.600	0.289	0.338
3	-	10122.493	10112.300	-	7.939	7.918

Let us now consider SOS-based Lasserre relaxations for problem (6.7). Using GloptiPoly3 as an interface to MOSEK, the we obtain an exact solution at the hierarchy's lowest level and the necessary semidefinite program is solved in only 0.23 seconds. Since the fastest exact conditional SAGE relaxation took 0.588 seconds, the SOS approach is the clear winner here. However, as a final comparison we consider another SAGE relaxation, here with $X = \mathbb{R}^n$, $G = G_{\text{all}}$, the natural ring, and $d = 3$. This also solves (6.7) exactly and the relative entropy program needs only 0.27 seconds to solve. Therefore while (6.6) comes with guarantees for compact X, it still performs well with noncompact X on this low-degree polynomial optimization problem.

6.4.2 Problem scaling

When signomial programs are considered in variables $t = \exp x$, the optimal solutions often have different decision variables span several orders of magnitude. One reason for this is that signomial models involve physical quantities with particular choices for units. Although it is *possible* to choose units where decision variables are similarly scaled at optimality, this may not be a natural thing to do from a modeling standpoint. This creates a need for algorithmic tools for signomial optimization that

are insensitive to scaling of the variable t. The following proposition shows that our hierarchy has such scale invariance.

Proposition 6.4.3. *Consider a signomial objective function f and a set of constraint signomials $G \cup G'$ where $\mathsf{X} = \{x \; : \; g(x) \geq 0 \text{ for all } g \in G'\}$ is convex. Given a vector $b \in \mathbb{R}^n$, construct translated problem data*

- *f_b defined by $f_b(x) = f(x - b)$,*

- *$G_b = \{x \mapsto g(x - b) \; : \; g \in G\}$, and*

- *$\mathsf{X}_b = \{x \; : \; g(x - b) \geq 0 \text{ for all } g \in G'\}$.*

Then the for every d and every signomial ring $\mathbb{R}[\mathcal{A}]$, the $\mathcal{A}$-degree d SAGE bound for Problem 6.5 is the same for problem data (f, G, X) and (f_b, G_b, X_b).

We should emphasize that scale invariance of a SAGE bound does not mean that the behavior of algorithms for REP are fully scale-invariant. Changes to problem scaling in finite precision arithmetic can affect both the speed at which an REP solver converges and even whether the solver converges at all. This proposition really shows that we are free to choose a coordinate system that works well for an REP solver without fear of changing the SAGE bound.

Proposition 6.4.3. Let h be a signomial on $\mathbb{R}^n$ and consider h_b defined by $h_b(x) = h(x - b)$. It is easy to verify that h is X-SAGE if and only if h_b is $[\mathsf{X} + b]$-SAGE. Additionally, it is clear that $\mathrm{supp}(h) = \mathrm{supp}(h_b)$, and this implies both $\deg_{\mathcal{A}}(h) = \deg_{\mathcal{A}}(h_b)$ and $\mathrm{supp}_{\mathcal{A}_d}^{-1}(h) = \mathrm{supp}_{\mathcal{A}_d}^{-1}(h_b)$. Finally, observe that $\mathsf{X}_b = \mathsf{X} + b$.

Using these facts we can map any feasible solution to Problem 6.6 for data (f, G, X) to a feasible solution to the analogous problem for data (f_b, G_b, X_b) without changing γ. By symmetry (essentially replacing b by $-b$) any solution to Problem 6.6 for data (f_b, G_b, X_b) can likewise be mapped to a feasible solution for problem data (f, G, X) without changing γ. As the set of feasible choices for γ is the same under these two formulations, we have that the $\mathcal{A}$-degree d SAGE bounds coincide. □

6.4.3 A benchmark in alkylation process design

Our next problem captures the design of an alkylation process in chemical engineering. A detailed derivation of the original model (in ten variables) may be found in [163] and the now-standard formulation (in seven variables) can be found in [41,

p. 7.2.1]. We have cleared denominators in the signomial program [41, p. 7.2.1] to obtain a cubic polynomial optimization problem. All coefficients in the formulation are positive; we refer the reader to [41] for their precise values.

$$\min_{t \in \mathbb{R}^7_{++}} c_1 t_1 + c_2 t_1 t_6 + c_3 t_3 + c_4 t_2 + c_5 - c_6 t_3 t_5 \tag{6.8}$$

$$\text{s.t. } t_1 - c_7 t_1 t_6^2 - c_8 t_3 + c_9 t_1 t_6 \geq 0$$

$$t_3 - c_{10} t_1 - c_{11} t_1 t_6 + c_{12} t_1 t_6^2 \geq 0$$

$$1 - c_{13} t_6^2 - c_{14} t_5 + c_{15} t_4 + c_{16} t_6 \geq 0$$

$$t_5 - c_{17} - c_{18} t_6 - c_{19} t_4 + c_{20} t_6^2 \geq 0$$

$$t_3 t_4 - c_{21} t_3 t_4 t_7 - c_{22} t_2 + c_{23} t_2 t_4 \geq 0$$

$$t_3 t_4 t_7 - c_{24} t_3 t_4 - c_{25} t_2 t_4 + c_{26} t_2 \geq 0$$

$$t_5 - c_{27} - c_{28} t_7 \geq 0 \qquad 1 - c_{29} t_5 + c_{30} t_7 \geq 0$$

$$1 - c_{31} t_3 + c_{32} t_1 \geq 0 \qquad t_3 - c_{33} t_1 - c_{34} \geq 0$$

$$t_3 t_4 - c_{35} t_2 + c_{36} t_2 t_4 \geq 0 \qquad t_2 - c_{37} t_2 t_4 - c_{38} t_3 t_4 \geq 0$$

$$1 - c_{39} t_1 t_6 - c_{40} t_1 + c_{41} t_3 \geq 0$$

$$t_1 - c_{42} t_3 - c_{43} + c_{44} t_1 t_6 \geq 0$$

$$(2000, 120, 3500, 93, 95, 12, 162) - t \geq \mathbf{0}$$

$$t - (1500, 1, 3000, 85, 90, 3, 145) \geq \mathbf{0}.$$

In order to reflect the polynomial structure in (6.8) we set $\mathcal{A} = \{\mathbf{0}, \delta_1, \ldots, \delta_7\}$. We approach this problem first through conditional SAGE. Let X be the set defined by the seventeen signomial inequalities $g(x) \geq 0$ where g has exactly one positive term, and have G include all twenty-eight constraints. Table 6.3 shows the results of using MOSEK 9.2 to compute $f_{\mathsf{K}}^{(3)}$ and $f_{\mathsf{K}}^{(4)}$ for (6.8). The reported solutions are feasible up to relative error at most 10^{-8}, however the primal solutions exhibit substantial absolute constraint violation.

The solution quality data in Table 6.3 lead us to speculate that (i) the reported objective is larger than actually possible for a feasible primal solution, and (ii) the given dual solution is actually suboptimal.

We can validate this speculation by considering a rescaled version of the problem. Specifically, we consider the change of variables

$$\hat{t} \leftarrow (t_i u_i / 10)^7_{i=1} \quad \text{for} \quad u = (2000, 120, 3500, 93, 95, 12, 162).$$

Table 6.3: MOSEK solution summary for SAGE relaxations to problem (6.8). The primal and dual objectives agree to six significant digits. The four-tuples give a solution's ℓ_∞-norm, then violations of elementwise affine constraints, violations of bound constraints, and violations of conic constraints.

d	objective	time	primal soln. norm & viols.	dual soln. norm & viols.
3	1224.52	8.6	$(1 \cdot 10^8, 3 \cdot 10^{-2}, 3 \cdot 10^{-3}, 0)$	$(3 \cdot 10^{10}, 0, 8 \cdot 10^{-6}, 0)$
4	1285.98	104.5	$(8 \cdot 10^7, 2 \cdot 10^{-1}, 1 \cdot 10^{-1}, 0)$	$(1 \cdot 10^{10}, 0, 2 \cdot 10^{-6}, 0)$

In terms of exponential-form signomials, this amounts to an affine shift $\hat{x} \leftarrow x - \log(10/u)$. Table 6.4 provides bounds and solver run-times for this rescaled problem using both X-SAGE (conditional SAGE) and $\mathbb{R}^n$-SAGE (ordinary SAGE). Applying solution recovery to the dual X-SAGE relaxation with $d = 4$ produces a de-scaled solution

$$t^\star \approx (1698.192,\ 53.662,\ 3031.305,\ 90.109,\ 95.000,\ 10.500,\ 153.540)$$

that is feasible within absolute error $3 \cdot 10^{-7}$ and satisfies $f(\log t^\star) \approx 1227.23$. The matching X-SAGE and $\mathbb{R}^n$-SAGE bounds show this is essentially optimal. Note that in contrast to the runtime performance of X-SAGE and $\mathbb{R}^n$-SAGE relaxations for problem (6.7), here the X-SAGE approach solves problem (6.8) over eight times faster than the $\mathbb{R}^n$-SAGE approach.

Table 6.4: SAGE bounds and solve times for the rescaled version of problem (6.8).

d	$\mathcal{A}$-degree d SAGE bound		solver runtime	
	conditional	ordinary	conditional	ordinary
3	1206.86	1125.12	3.83	2.83
4	1227.23	1224.98	50.56	84.12
5	-	1227.23	-	429.43

6.4.4 Design of a chemical reactor system

Here we consider the design of a chemical reactor system as described by Blau and Wilde in [164] and [39]. This problem is a proper signomial program and we approach it through the naive ring. None of the constraints in this problem are convex in x, however we can infer convex constraints by considering the case $g(x) \geq 0$ in each of the constraints $g(x) = 0$. We apply the hierarchy (6.6) to this problem by taking X as the convex set cut out by these five inequality constraints.

$$\min_{t\in\mathbb{R}^8_{++}} \quad 2.0425\, t_1^{0.782} + 52.25\, t_2 + 192.85\, t_2^{0.9} + 5.25\, t_2^3 + 61.465\, t_6^{0.467} \tag{6.9}$$
$$+ 0.01748\, t_3^{1.33}/t_4^{0.8} + 100.7\, t_4^{0.546} + 3.66\cdot 10^{-10}\, t_3^{2.85}/t_4^{1.7}$$
$$+ 0.00945\, t_5 + 1.06\cdot 10^{-10}\, t_5^{2.8}/t_4^{1.8} + 116\, t_6 - 205\, t_6 t_7 - 278\, t_2^3 t_7$$

$$\text{s.t.} \quad 1 - 129.4/t_2^3 - 105/t_6 = 0$$
$$1 - 1.03\cdot 10^5\, t_2^3 t_7/(t_3 t_8) - 1.2\cdot 10^6/(t_3 t_8) = 0$$
$$1 - 4.68\, t_2^3/t_1 - 61.3\, t_2^2/t_1 - 160.5\, t_2/t_1 = 0$$
$$1 - 1.79\, t_7 - 3.02\, t_2^3 t_7/t_6 - 35.7\,/t_6 - 1 = 0$$
$$1 - 1.22\cdot 10^{-3}\, t_3 t_8/(t_4^{0.2} t_5^{0.8}) - 1.67\cdot 10^{-3}\, t_8 t_3^{0.4}/t_4^{0.43}$$
$$- 3.6\cdot 10^{-5}\, t_3 t_8/t_4 - 2\cdot 10^{-3}\, t_3 t_8/t_5 - 4\cdot 10^{-3}\, t_8 = 0$$

This initial problem statement is horrifically scaled – the coefficients in the objective alone span twelve orders of magnitude. Trying to solve even the lowest-level SAGE relaxations with MOSEK returns "unknown" status codes here. We therefore scale the variables about the initial estimates provided in [39]

$$\tilde{t} = (10^3,\ 10,\ 10^5,\ 10^2,\ 10^5,\ 10^3,\ 10^{-1}, 10)$$

and we call solvers with a scaled objective $\hat{f} := f/10^4$.

The coefficients in the scaled problem span only four orders of magnitude and the SAGE relaxations can be solved reliably. We compute

$$f_{\mathsf{K}}^{(1)} = 16377.32 \quad \text{in} \quad 0.13 \ \text{seconds, and}$$
$$f_{\mathsf{K}}^{(2)} = 17462.73 \quad \text{in} \quad 24.37 \ \text{seconds.}$$

We run solution recovery on the dual formulation for $f_{\mathsf{K}}^{(2)}$ to obtain a point x', and refine this with COBYLA (a zeroth-order local solver, see [165]) to get x''. These solutions satisfy

$$f(x') = 17486.52 \quad \text{and} \quad \|G(x')\|_\infty = 2.05\cdot 10^{-5},\ \text{as well as}$$
$$f(x'') = 17485.99 \quad \text{and} \quad \|G(x'')\|_\infty = 5.85\cdot 10^{-15}$$

where we have abused notation by writing $G(x) := (g(x))_{g\in G}$. The point x'' is feasible to nearly machine precision and so we can reasonably conclude $f(x'') \geq f_{\mathsf{K}}^\star$. We combine this with the SAGE bound to obtain $(f_{\mathsf{K}}^\star - f_{\mathsf{K}}^{(2)})/f_{\mathsf{K}}^\star \leq 0.0013$. That is, the $\mathcal{A}$-degree 2 SAGE relaxation solves (6.9) within one percent relative error.

One can alternatively approach this problem through a global solver from the traditional nonlinear programming community. We tested BARON, ANTIGONE, LINDO, and SCIP – which together are four out of the five global nonlinear solvers in the Mittelmann benchmarks.[1] We ran each of these solvers by passing it (6.9) once in variables t and once in variables x. When passing the problem in variables t we had to disable warnings from GAMS about unbounded monomials with negative exponents. For all configurations we used a time limit of 7200 seconds, allocated 8 threads, and left the machine otherwise unused.

In these experiments, SCIP terminated after 7200 seconds with no feasible solution and no lower bound. Precise results for the remaining solvers are reported in Table 6.4.4. The overall takeaway is that SAGE produced the same solution as these solvers, but with an REP that could be solved in half the time as the fastest of these methods. Only LINDO was able to certify its solution as globally optimal. By contrast with LINDO, the performance of SAGE is independent of whether signomials are considered as generalized polynomials in t or as functions of x.

Table 6.5: Results of applying global solvers from GAMS to a reactor design problem in chemical engineering (6.8). All solvers returned a solution with objective value approximately equal to 17485.99.

	Using t as optimization variable		Using x as optimization variable	
	solver time (s)	lower bound	solver time (s)	lower bound
BARON	163	$-\infty$	7200	$-\infty$
ANTIGONE	145	-16880.380	7200	$-\infty$
LINDO	1468	17484.314	50	17485.988

Remark 6.4.4. This model first appeared in [164], which was written very much for practicing engineers. Later, the problem was considered as an example for a proposed algorithm for equality-constrained signomial programming [39]. We used the formulation [39] since it was easier to read than that in [164]. However, there is a clear typo in [39, Equation 4]: the term "d_V^2/W" appears with two different coefficients. We consulted the original paper [164] and believe the correct version of [39, Equation 4] is $4.68d_V^3/W + 6.13d_V^2/W + 160.5d_V/W = 1$. The bounds and runtime results reported here are equally informative regardless or whether or correction was valid.

---

[1]The fifth solver (COUENNE) was not available in our version of GAMS (33.2.0).

6.5 Outer-approximations and signomial moment theory

In this section we establish a theory of moments for signomials. We begin with definitions that are analogous to those in the moment-SOS literature for polynomial optimization. We then state our main result – a method to develop successively stronger outer-approximations for the cone of signomials in $\mathbb{R}[\mathcal{A}]$ that are nonnegative on K. Proving our main result requires establishing a signomial Riesz-Haviland theorem and a moment-determinacy result. We draw liberally from [123] in its exposition on analogous results for polynomials and from [112] for basic ingredients proven for abstract algebras.

6.5.1 Definitions

Throughout this section and the next, we use $\mathcal{A}_\infty$ to denote the smallest subset of $\mathbb{R}^n$ that contains all the $\mathcal{A}_d$. Clearly, signomials $f \in \mathbb{R}[\mathcal{A}]$ are in one-to-one correspondence with finitely supported sequences $\boldsymbol{f} = (f_\alpha)_{\alpha \in \mathcal{A}_\infty}$. Most of our arguments in this section focus on the dual space to $\mathbb{R}[\mathcal{A}]$, which we identify with $\mathbb{R}^{\mathcal{A}_\infty}$. Sequences $\boldsymbol{y} \in \mathbb{R}^{\mathcal{A}_\infty}$ are associated to linear functions $L_{\boldsymbol{y}} : \mathbb{R}[\mathcal{A}] \to \mathbb{R}$ defined by $L_{\boldsymbol{y}}(\mathrm{e}^\alpha) = y_\alpha$. We call $L_{\boldsymbol{y}}$ the *Riesz functional* of $\boldsymbol{y}$.

We take note of two ways to express the output of a Riesz functional. In full generality, applying $L_{\boldsymbol{y}}$ to $f = \sum_{\alpha \in \mathcal{A}_\infty} f_\alpha \mathrm{e}^\alpha$ can be written as $L_{\boldsymbol{y}}(f) = \langle \boldsymbol{f}, \boldsymbol{y} \rangle$. We are more interested in the special case when $\boldsymbol{y}$ is a *moment sequence*. That is, when there is a finite Borel measure μ for which

$$y_\alpha = \int \mathrm{e}^\alpha(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x}) \quad \text{for all} \quad \boldsymbol{\alpha} \in \mathcal{A}_\infty.$$

We call such a μ a *representing measure* for $\boldsymbol{y}$. Moment sequences are important because the value of the Riesz functional can be expressed as $L_{\boldsymbol{y}}(f) = \int f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$.

Next, we introduce a concept directly analogous to the "localizing matrix" in the moment-SOS literature. In our case, the *localizer* induced by a sequence $\boldsymbol{y} \in \mathbb{R}^{\mathcal{A}_\infty}$ and a dimensional parameter $d \in \mathbb{Z}_{++} \cup \{\infty\}$ is the linear operator

$$V_d(\,\cdot\,\boldsymbol{y}) \colon \mathbb{R}[\mathcal{A}] \to \mathbb{R}^{\mathcal{A}_d} \quad \text{defined by} \quad f \mapsto (L_{\boldsymbol{y}}(f\mathrm{e}^\alpha) \,:\, \boldsymbol{\alpha} \in \mathcal{A}_d).$$

We abbreviate the case $V_d(\mathrm{e}^{\boldsymbol{0}}\boldsymbol{y})$ by $V_d(\boldsymbol{y})$.

Localizers help us make abstract arguments concrete. For example, a localizer can truncate infinite sequences $\boldsymbol{y} \in \mathbb{R}^{\mathcal{A}_\infty}$ to $V_d(\boldsymbol{y}) = (y_\alpha \,:\, \boldsymbol{\alpha} \in \mathcal{A}_d)$. One can also see that if the $\mathcal{A}$-degree of a signomial f is at most d, then we can identify f by a vector of coefficients $\boldsymbol{f} \in \mathbb{R}^{\mathcal{A}_d}$ and evaluate a Riesz functional by $L_{\boldsymbol{y}}(f) = \langle \boldsymbol{f}, V_d(\boldsymbol{y}) \rangle$.

That second point is important: given a convex cone $C \subset \mathbb{R}[\mathcal{A}]_{\leq d}$, the condition that $V_d(\mathbf{y}) \in C^\dagger$ is the same as $L_{\mathbf{y}}(f) \geq 0 \forall f \in C$.

6.5.2 Main result: nonnegativity cones from the outside

Theorem 6.5.1. *Let $\mathcal{A}$ be injective and K be compact. Consider any sequence $\mathscr{C} = (C_d)_{d \geq 1}$ of closed convex cones where* (i) $C_d \subset \mathbb{R}[\mathcal{A}]_{\leq d}$, (ii) *every $f \in C_d$ is K-nonnegative, and* (iii) *for every K-positive $f \in \mathbb{R}[\mathcal{A}]$, there exists some d for which $f \in C_d$. If $\mathbf{y}$ is the moment sequence of a Borel measure μ with $\operatorname{supp} \mu = K$, then $f \in \mathbb{R}[\mathcal{A}]$ is K-nonnegative if and only if $V_d(f\mathbf{y}) \in C_d^\dagger$ for all integers $d \geq 1$.*

We believe Theorem 6.5.1 is interesting as-stated, but it is only useful if we have access to an appropriate sequence of sets $\mathscr{C}$. Such sequences can nominally be obtained through our Positivstellensatz in Theorem 6.3.1. However, as we saw in Section 6.4, grading certificates provided by our Positivstellensatz according to $\mathcal{A}$-degree is complicated. We therefore instead focus on when $K = X$ for a compact convex set X and rely on Corollary 6.3.2 to obtain the sequence $\mathscr{C}$. Specifically, for $d \in \mathbb{Z}_{++} \cup \{\infty\}$, we define nested cones

$$C_X^d(\mathcal{A}) = \left\{ f \in \mathbb{R}[\mathcal{A}]_{\leq d} \ : \ \exists r \in \mathbb{N} \text{ with } \left(\sum_{\alpha \in \mathcal{A}} e^\alpha \right)^r f \in C_X(\mathcal{A}_d) \right\}. \quad (6.10)$$

The sets $C_X^d(\mathcal{A})$ consist of X-SAGE signomials supported on $\mathcal{A}_d$, as well as many signomials that require modulation before admitting a SAGE certificate. Observe that in particular $C_X^1(\mathcal{A}) = C_X(\mathcal{A}_1)$ and that when X is compact $C_X^\infty(\mathcal{A})$ contains all $f \in \mathbb{R}[\mathcal{A}]$ that are positive on X.

Corollary 6.5.2. *Let $\mathcal{A}$ be injective, X be a compact convex set, and $\mathbf{y}$ be the moment sequence induced by a Borel measure μ with support X. A signomial $f \in \mathbb{R}[\mathcal{A}]$ is X-nonnegative if and only if $V_d(f\mathbf{y}) \in C_X^d(\mathcal{A})^\dagger$ for all integers $d \geq 1$.*

Corollary 6.5.2 is used in Section 6.6 to prove of a hierarchy of REP relaxations for approaching $f_X^\star$ from above. We can develop another consequence of Theorem 6.5.1 by considering how these cones $C_X^d(\mathcal{A})$ are nested. Specifically, we can obtain nested *outer approximations* for cones of X-nonnegative signomials of given $\mathcal{A}$-degree.

Corollary 6.5.3. *Let $\mathcal{A}$ be injective, X be a compact convex set, and $\mathbf{y}$ be the moment sequence induced by a Borel measure μ with support X. Fix an integer $d \geq 1$, let $P_X(\mathcal{A})$ denote the cone of X-nonnegative signomials in $\mathbb{R}[\mathcal{A}]_{\leq d}$, and define*

$$Q_X^\ell(\mathcal{A}) = \left\{ f \in \mathbb{R}[\mathcal{A}]_{\leq d} \ : \ V_\ell(f\mathbf{y}) \in C_X^\ell(\mathcal{A})^\dagger \right\} \quad \text{for} \quad \ell \in \mathbb{Z}_{++} \cup \{\infty\}.$$

We have $Q_X^1(\mathcal{A}) \supset Q_X^2(\mathcal{A}) \supset \cdots \supset Q_X^\infty(\mathcal{A}) = P_X(\mathcal{A})$.

6.5.3 Supporting results for signomial moments

Here we present two basic results for signomial moments that go into our proof of Theorem 6.5.1. The first of these results is a Riesz-Haviland type theorem, which describes when the condition "$L_y(f) \geq 0$ for all K-nonnegative f" ensures that y has a representing measure supported on K. The second of these results concerns when that representing measure is unique. Our proofs results draw heavily from Marshall's book [112, §3].

Theorem 6.5.4. *Suppose $K \subset \mathbb{R}^n$ is closed and that for every unbounded sequence $(x_t)_{t \in \mathbb{N}} \subset K$, we have $\limsup_t \max_{\alpha \in \mathcal{A}} \langle \alpha, x_t \rangle = +\infty$. Given a sequence $y \in \mathbb{R}^{\mathcal{A}_\infty}$, there exists a finite Borel measure μ on K such that*

$$\int e^\alpha(x) \, d\mu(x) = y_\alpha \quad \text{for all} \quad \alpha \in \mathcal{A}_\infty$$

if and only if $L_y(f) \geq 0$ for all signomials $f \in \mathbb{R}[\mathcal{A}]$ nonnegative on K.

Clearly, Theorem 6.5.4 applies when K is compact. If K is convex, then the hypothesis is satisfied if and only if the intersection of the recession cone of $-\mathcal{A}K$ and $\mathbb{R}_+^{\mathcal{A}}$ consists only of the origin. If we consider two closed sets K, K' where $K \subset K'$ and $(\mathcal{A}, K')$ satisfy the hypothesis of Theorem 6.5.4, then $(\mathcal{A}, K)$ also satisfy the hypothesis of Theorem 6.5.4. In particular, we may conclude that if $\mathcal{A}$ contains the origin in the interior of its convex hull, then $(\mathcal{A}, K)$ satisfies the hypothesis of Theorem 6.5.4 for any $K \subset \mathbb{R}^n$.

To prove Theorem 6.5.4 we rely on the following result. Let $\text{Cont}(\Omega, \mathbb{R})$ denote the ring ($\mathbb{R}$-algebra) of all continuous functions $f \colon \Omega \to \mathbb{R}$.

Theorem 6.5.5 (Theorem 3.2.2, [112]). *Let A be an $\mathbb{R}$-algebra, Ω a Hausdorff space, and $\hat{}\colon A \to \text{Cont}(\Omega, \mathbb{R})$ an $\mathbb{R}$-algebra homomorphism. Assume there exists a $p \in A$ such that $\hat{p} \geq 0$ in Ω and, for each integer $k \geq 1$, the sublevel set $\Omega_k = \{x \in \Omega : \hat{p}(x) \leq k\}$ is compact. Then, for any linear functional $L \colon A \to \mathbb{R}$ satisfying $L(\{a \in A : \hat{a} \geq 0 \text{ on } \Omega\}) \subset \mathbb{R}_+$, there exists a Borel measure μ on Ω such that $L(a) = \int_\Omega \hat{a} \, d\mu$ for all $a \in A$.*

Proof of Theorem 6.5.4. The claim follows from Theorem 6.5.5 by the following identification: $A = \mathbb{R}[\mathcal{A}], \Omega = K$, and $\hat{}\colon \mathbb{R}[\mathcal{A}] \to \text{Cont}(K, \mathbb{R})$ with $\hat{f}(x) = f(x)$ for all $x \in K$, i.e., $\hat{f}$ is the restriction of the signomial f to K. Consider the

distinguished signomial $\psi = \sum_{\alpha \in \mathcal{A}} e^{\alpha}$. It suffices to show that for any k the sublevel set $\Omega_k = \{x \in \mathsf{K} : \hat{\psi}(x) \leq k\}$ is compact. It is obvious that Ω_k is closed. Moreover, in order for a point x to belong to Ω_k it is necessary that $\langle \alpha, x \rangle \leq \log k$ for all $\alpha \in \mathcal{A}$. By the theorem's assumption, any unbounded sequence in K cannot satisfy this property. Therefore all sequences $(x_t)_{t \in \mathbb{N}} \subset \Omega_k$ are bounded, which implies compactness of Ω_k. □

The proof of this section's main result requires that moment sequences admit unique representing measures. A measure is called *moment determinate* if it is the unique Borel measure that gives rise to its moment sequence. It is well known that in the polynomial case, measures supported on compact sets are moment-determinate. We prove the same is true for signomials.

Theorem 6.5.6. *Suppose $\mathcal{A}$ is injective and K is compact. If y is a moment sequence of two finite Borel measures μ_1, μ_2 with* $\text{supp}\,\mu_1 \subset \mathsf{K}$ *and* $\text{supp}\,\mu_2 \subset \mathsf{K}$, *then $\mu_1 = \mu_2$.*

Proof. If we can show that $L_y : \mathbb{R}[\mathcal{A}] \to \mathbb{R}$ has a unique continuous extension $\overline{L}_y : \text{Cont}(\mathsf{K}, \mathbb{R}) \to \mathbb{R}$, then the theorem's claim will follow from the Riesz Representation Theorem (see [112, §3.2.1]). Following [112], the uniqueness of such an extension can be stated as follows: for every $\phi \in \text{Cont}(\mathsf{K}, \mathbb{R})$, we have

$$\sup_{f \in \mathbb{R}[\mathcal{A}]} \{L_y(f) : \phi - f \geq 0 \text{ on } \mathsf{K}\} = \inf_{f \in \mathbb{R}[\mathcal{A}]} \{L_y(f) : f - \phi \geq 0 \text{ on } \mathsf{K}\}. \quad (6.11)$$

It is easily shown that (6.11) holds if every function in $\text{Cont}(\mathsf{K}, \mathbb{R})$ can be approximated to arbitrary precision (in sup norm) by a signomial in $\mathbb{R}[\mathcal{A}]$. The Stone-Weierstrass Theorem tells us that such an approximation exists if signomials in $\mathbb{R}[\mathcal{A}]$ can separate points, i.e., if for every pair of distinct $x, x' \in \mathsf{K}$, there exists an $f \in \mathbb{R}[\mathcal{A}]$ for which $f(x) \neq f(x')$.

We now show that signomials in $\mathbb{R}[\mathcal{A}]$ can separate points. Let x and x' be distinct points in $\mathbb{R}^n$. By the injectivity of $\mathcal{A}$, the images $z := \mathcal{A}x$ and $z' := \mathcal{A}x'$ are likewise distinct in $\mathbb{R}^{\mathcal{A}}$. Recall that these vectors have components $z_\alpha = \langle \alpha, x \rangle$ and $z'_\alpha = \langle \alpha, x' \rangle$, so the condition that $z \neq z'$ means there exists a $\beta \in \mathcal{A}$ for which $\langle \beta, x \rangle \neq \langle \beta, x' \rangle$. We exponentiate both sides of that non-equality to find

$$e^{\beta}(x) = \exp\langle \beta, x \rangle \neq \exp\langle \beta, x' \rangle = e^{\beta}(x'),$$

which that tells us that $e^{\beta} \in \mathbb{R}[\mathcal{A}]$ separates x, x'. □

It is of note that in the polynomial case, there are conditions under which we can conclude that a moment sequence is generated by a moment-determinate measure with *noncompact support*. See, for example, [123, Proposition 2.37(a)]

Remark 6.5.7. Our proof of Theorem 6.5.4 somewhat informally called $\Omega = \mathsf{K}$ a Hausdorff space. If $\mathsf{K} \subset \mathbb{R}^n$ were open, then we could formally take the Hausdorff space to be all open subsets of K using the standard topology on $\mathbb{R}^n$. However, Theorem 6.5.4 clearly allows for sets K that are not open. So, what is the actual topology used to create a Hausdorff space from closed subsets of $\mathbb{R}^n$? Marshall glosses over this, but repeatedly indicates that it is valid to identify $\Omega = \mathsf{K}$ when K is any closed subset of $\mathbb{R}^n$. The differences between Marshall's treatment of the polynomial case and our proof can be resolved by requiring that $\operatorname{conv}(\mathcal{A})$ contains the origin in its interior.

6.5.4 Proof of Theorem 6.5.1

For this proof we denote the cone of K-nonnegative signomials of $\mathcal{A}$-degree at most d by P_d. Properties (i) and (ii) of $\mathscr{C}$ tell us that $\mathsf{C}_d \subset \mathsf{P}_d$.

Suppose f is K-nonnegative. We will show that $V_d(f\boldsymbol{y}) \in \mathsf{C}_d^\dagger$ holds for all d. As a first step, define $\hat{\boldsymbol{y}} \in \mathbb{R}^{\mathcal{A}_\infty}$ by $\hat{y}_\alpha = \int \mathrm{e}^\alpha(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$ for all $\alpha \in \mathcal{A}_\infty$. Because f is K-nonnegative, the differential quantity $\mathrm{d}\phi(\boldsymbol{x}) = f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$ defines a finite Borel measure on K, so $\hat{\boldsymbol{y}}$ is a moment sequence. Meanwhile, the simple identity $L_{\boldsymbol{y}}(f\mathrm{e}^\alpha) = \hat{y}_\alpha$ tells us that $V_d(f\boldsymbol{y}) = V_d(\hat{\boldsymbol{y}})$ for all d. Combine these to see that $V_d(f\boldsymbol{y}) \in \mathsf{P}_d^\dagger$ for all d. The result follows since $\mathsf{P}_d^\dagger \subset \mathsf{C}_d^\dagger$.

Now we address the theorem's other claim: we show that if $V_d(f\boldsymbol{y}) \in \mathsf{C}_d^\dagger$ for all d, then f is nonnegative on K.

Once again, we define $\hat{\boldsymbol{y}} \in \mathbb{R}^{\mathcal{A}_\infty}$ by $\hat{y}_\alpha = \int \mathrm{e}^\alpha(\boldsymbol{x}) f(\boldsymbol{x}) \, \mathrm{d}\mu(\boldsymbol{x})$ so that $V_d(\hat{\boldsymbol{y}}) = V_d(f\boldsymbol{y})$. Let d be any fixed positive integer. We claim that $L_{\hat{\boldsymbol{y}}}(g) \geq 0$ for all $g \in \mathsf{P}_d$; by a continuity argument this claim holds if $L_{\hat{\boldsymbol{y}}}(g) \geq 0$ for all K-positive g with $\deg_{\mathcal{A}}(g) \leq d$. Let us fix such a g. By property (iii) of $\mathscr{C}$, there exists an integer $d' \geq \deg_{\mathcal{A}}(g)$ for which $g \in \mathsf{C}_{d'}$. Now, our assumption on $\hat{\boldsymbol{y}}$ includes $V_{d'}(\hat{\boldsymbol{y}}) \in \mathsf{C}_{d'}^\dagger$, which tells us $L_{\hat{\boldsymbol{y}}}(g) \geq 0$! Therefore $L_{\hat{\boldsymbol{y}}}(g) \geq 0$ for every $g \in \mathsf{P}_d$ for our *arbitrary* fixed d. We can now invoke Theorem 6.5.4 to see that there is some Borel measure ψ with $\operatorname{supp} \psi \subset \mathsf{K}$ and $\hat{\boldsymbol{y}}$ as its moment sequence. Because K is compact and $\mathcal{A}$ is injective, Theorem 6.5.6 tells us that ψ is unique.

By now we have shown that there is a unique Borel measure ψ for which

$$\int e^{\alpha}(x) f(x) \, d\mu(x) = \int e^{\alpha}(x) \, d\psi(x) \quad \text{for all} \quad \alpha \quad \text{in} \quad \mathcal{A}_{\infty}. \quad (6.12)$$

Getting from (6.12) to "$f \geq 0$ on K" requires two steps. The main step is to carefully use moment determinacy (Theorem 6.5.6) to show that $f(x) \, d\mu(x)$ is a Borel measure on K. The claim then follows by an application of [123, Lemma 3.1].

So we turn to showing that $f(x) \, d\mu(x)$ induces a Borel measure. Begin by introducing $\mathsf{B}_1 = \{x \in \mathsf{K} : f(x) \geq 0\}$ and $\mathsf{B}_2 = \{x \in \mathsf{K} : f(x) < 0\}$. We want to show that B_2 is empty but we have no tools to do this directly. Instead, we use $\mathsf{B}_1, \mathsf{B}_2$ to define the functions

$$\phi_1(\mathsf{B}) = \int_{\mathsf{B} \cap \mathsf{B}_1} f(x) \, d\mu(x) \quad \text{and} \quad \phi_2(\mathsf{B}) = \int_{\mathsf{B} \cap \mathsf{B}_2} (-f(x)) \, d\mu(x).$$

These functions are finite Borel measures since f is continuous and K is compact. We can therefore define the *signed measure* $\phi = \phi_1 - \phi_2$ and note that $\int e^{\alpha}(x) \, d\phi(x) = \int e^{\alpha}(x) \, d\psi(x)$ for all $\alpha \in \mathcal{A}_{\infty}$ – equations that can be rewritten as

$$\int e^{\alpha}(x) \, d\phi_1(x) = \int e^{\alpha}(x) \, d(\psi + \phi_2)(x) \quad \text{for all } \alpha \quad \text{in} \quad \mathcal{A}_{\infty}.$$

The key is that now, ϕ_1 and $\psi + \phi_2$ are Borel measures, therefore the fact that their moments match lets us use Theorem 6.5.6 to conclude that they are unique, i.e., $\phi_1 = \psi + \phi_2$. From here we simply rewrite $\phi = \phi_1 - \phi_2 = \psi$ to see that since ψ is a Borel measure, so is ϕ.

6.6 A complete hierarchy of upper bounds

In this section we develop a signomial analog to hierarchies of upper bounds for polynomial minimization. The original idea for this approach comes from [143]. We were made aware of this idea in a more abstract sense through a presentation by de Klerk at the 2019 ICCOPT meeting in Berlin, Germany (see [153, 156]).

Throughout this section we consider the problem of computing

$$f_{\mathsf{X}}^{\star} := \inf\{f(x) : x \in \mathsf{X}\}. \quad (6.13)$$

We assume that X is convex for ease of exposition; our proof techniques show that the deeper assumption is having access to a sequence of arbitrarily strong inner-approximations of a given signomial nonnegativity cone. We also use the notation where $y \in \mathbb{R}^{\mathcal{A}_{\infty}}$ is the moment sequence for a *reference measure* μ supported on X.

6.6.1 A simple upper bound

Let ψ be a signomial in $C_X(\mathcal{A}_1)$ with coefficient vector $\boldsymbol{\psi} \in \mathbb{R}^{\mathcal{A}_1}$. If we assume that $L_y(\psi) = 1$ (equivalently, $\langle V_1(y), \boldsymbol{\psi} \rangle = 1$), then the differential quantity $\mathrm{d}\psi(x) := \langle \boldsymbol{\psi}, \exp \mathcal{A}x \rangle \, \mathrm{d}\mu(x)$ defines a probability distribution on X. For any signomial $f \in \mathbb{R}[\mathcal{A}]$, direct calculations yield

$$
\inf\{f(x) \,:\, x \in X\} \leq \int f(x) \, \mathrm{d}\psi(x)
$$
$$
= \sum_{\alpha \in \mathcal{A}_1} \psi_\alpha L_y(f \mathrm{e}^\alpha)
$$
$$
= \langle \boldsymbol{\psi}, V_1(fy) \rangle.
$$

Therefore, if we can compute the localizers $V_1(fy), V_1(y) \in \mathbb{R}^{\mathcal{A}_1}$, then we can solve the tractable convex program

$$
\lambda_1 := \inf \{ \langle V_1(fy), \boldsymbol{\psi} \rangle \,:\, \langle V_1(y), \boldsymbol{\psi} \rangle = 1, \ \psi \in C_X(\mathcal{A}_1) \} \tag{6.14}
$$

to obtain an upper bound $\lambda_1 \geq f_X^\star$. In practice it is very important to ask how to construct the vectors $V_1(fy), V_1(y)$ given the reference measure μ (or simply given the moment sequence y). We provide some remarks on how that can be done later in this section (see Subsection 6.6.2). Although, we will say up front that this section's results are of only theoretical interest at the moment.

6.6.2 The hierarchy of upper bounds

Consider the following optimization problems, parameterized by integers $\ell \geq 1$:

$$
\lambda_\ell = \inf_{\boldsymbol{\psi}} \{ \langle V_\ell(fy), \boldsymbol{\psi} \rangle \,:\, \langle V_\ell(y), \boldsymbol{\psi} \rangle = 1, \ \psi \in C_X^\ell(\mathcal{A}) \}. \tag{6.15}
$$

One can show that $\lambda_\ell \geq f_X^\star$ by similar reasoning as before. In order to prove convergence $\lambda_\ell \to f_X^\star$, we need the consider the dual to (6.15). This dual is given by

$$
\lambda_\ell = \sup_\lambda \{ \lambda \in \mathbb{R} \,:\, V_\ell(fy) - \lambda V_\ell(y) \in C_X^\ell(\mathcal{A})^\dagger \}. \tag{6.16}
$$

We prove strong duality for (6.15)-(6.16) towards the end of this section, along with some other exploration of the special structure in problem (6.16).

Theorem 6.6.1. *Let $\mathcal{A}$ be injective, f be a signomial in $\mathbb{R}[\mathcal{A}]$, and $X \subset \mathbb{R}^n$ be a compact convex set. The hierarchy given by the sequence $(\lambda_\ell)_{\ell \geq 1}$ is complete, i.e., $\lambda_\ell \downarrow f_X^\star$ as $\ell \to \infty$.*

Proof. Assume for now that strong duality holds for (6.15)–(6.16). By consideration to the formulation in (6.15), it is easy to see that λ_ℓ is decreasing. Simply rewrite that problem as

$$\lambda_\ell = \inf_\psi \left\{ \int f\psi \, d\mu \; : \; \int \psi \, d\mu = 1, \, \psi \in C_X^\ell(\mathcal{A}) \right\}$$

and note that the size of the feasible set for this problem is increasing in ℓ.

Since X is compact and f is continuous, we know that $f_X^\star$ is a real number. The sequence $(\lambda_\ell)_{\ell \geq 1}$ is therefore decreasing and bounded below by $f_X^\star$. Let $\lambda^\star$ denote the limit of this sequence and suppose that $\lambda^\star > f_X^\star$. By consideration to the formulation (6.16), this limit satisfies

$$V_\ell((f - \lambda^\star)y) \in C_X^\ell(\mathcal{A}) \quad \text{for all} \quad \ell \in \mathbb{Z}_{++} \cup \{\infty\}.$$

But then by Corollary 6.5.2, we have that $f - \lambda^\star$ is nonnnegative on X! This contradicts our earlier assumption that $\lambda^\star > f_X^\star$, and so we must have $\lambda^\star = f_X^\star$, and this completes our proof. □

Note that in general the convergence is only asymptotic and not finite. That is, in general, there does not exist a finite ℓ such that $\lambda_\ell = f_X^\star$.

Theoretically the methodology developed above, particularly Theorem 6.6.1, can be applied to any signomial f and compact convex set X. However, for practical purposes we are limited to cases where we know or can actually compute the moments of the reference measure μ on X. There are a few interesting cases where the sequence of moments $y \in \mathbb{R}^{\mathcal{A}_\infty}$ can be derived either in closed form or numerically. An especially simple case is when μ is the uniform measure on a box X. Other examples with closed-form expressions for signomial moments include uniform measures over ellipsoids [166, Theorem 3.2] and solid simplices [166, Theorem 2.6]. If μ is the uniform measure on a polytope then one can nominally compute moments by triangulating that polytope with simplices (see [167]).

6.6.3 More on the one-variable "primal" relaxation

Although we introduced problem (6.15) before (6.16), we actually call (6.16) the *primal* in the primal-dual pair. We use this terminology because it is consistent with analogous hierarchies for polynomial optimization. We now establish strong duality for this primal-dual pair.

Proposition 6.6.2. *For any probability measure μ with* supp $\mu \subset X$ *(X not necessarily compact, or convex) and associated moment sequence $y \in \mathbb{R}^{\mathcal{A}_\infty}$, the primal-dual pair (6.16)-(6.15) exhibits strong duality. If $f_X^\star > -\infty$ and* supp μ *has nonempty interior, then (6.15) attains an optimal solution.*

Proof. First we prove strong duality in the sense of objective values. Since y is a moment sequence, we have that $V_\ell(y) = (y_\alpha : \alpha \in \mathcal{A}_\ell)$ is elementwise positive. It is obvious that the posynomial $\psi = \sum_{\alpha \in \mathcal{A}_\ell} e^\alpha$ belongs to the interior of $C_X^\ell(\mathcal{A})$, and so defining $s := \langle V_\ell(y), \psi \rangle > 0$, the signomial $\psi' = \psi/s$ is strictly feasible for the dual problem (6.15). The claim follows by invoking Slater's condition.

Now suppose $f_X^\star > -\infty$ and X has nonempty interior. Then for any $\lambda < f_X^\star - 1$ and every nonzero function $\psi \in C_X^\ell(\mathcal{A})$, we have

$$\langle V_\ell((f - \lambda)y), \psi \rangle = \int \underbrace{(f(x) - \lambda)}_{>1} \psi(x)\, \mathrm{d}\mu(x) > \int \psi(x)\, \mathrm{d}\mu(x) > 0.$$

By the above inequalities, we have that $V_\ell((f - \lambda)y) = V_\ell(fy) - \lambda V_\ell(y)$ belongs to the interior of $C_X^\ell(\mathcal{A})^\dagger$. Therefore (6.16) is strictly feasible, and by Slater's condition (6.15) attains an optimal solution. $\qquad\square$

For the remainder of this section we speak to nice structures in problem (6.16). Beyond the fact that the problem nominally has a single variable, if X is a polyhedron, then $C_X(\mathcal{A})^\dagger \subset \mathbb{R}^{\mathcal{A}}$ can be represented by finitely many power-cone inequalities without any lifting (apply Corollary 5.4.5 and Proposition 5.4.7). Therefore when X is a polyhedron it is nominally possible to write (6.16) as a finite power-cone program in a single variable. However, finding the lifting-free power-cone representation of $C_X(\mathcal{A})^\dagger$ is usually not tractable. In practice one should solve (6.16) using the standard relative entropy lift for $C_X(\mathcal{A})^\dagger$ and by bisection on λ.

Consider the special case of (6.16) with $\ell = 1$. Then for each fixed value of λ, we can solve $|\mathcal{A}_1|$ convex feasibility problems to determine if $V_1(fy) - \lambda V_1(y)$ belongs to $C_X(\mathcal{A}_1)^\dagger$. Each of these convex feasibility problems is in n variables, and if $X = \{x : Gx \le h\}$ is a polyhedron, then these feasibility problems are linear programs (as $z_\beta/v_\beta \in X$ is represented as $Gz_\beta \le hv_\beta$).

We now turn to finding a representation of $C_X^\ell(\mathcal{A})^\dagger$ with $\ell > 1$. For a positive integer j and a signomial $g \in \mathbb{R}[\mathcal{A}]_{\le i}$, the *moment reduction map* $M_j(g) : \mathbb{R}^{\mathcal{A}_{i+j}} \to \mathbb{R}^{\mathcal{A}_j}$ is the linear operator whose α^{th} row ($\alpha \in \mathcal{A}_j$) is the coefficient vector of $e^\alpha g$ expressed in $\mathbb{R}^{\mathcal{A}_{i+j}}$.

Lemma 6.6.3. *Let g be a signomial in* $\mathbb{R}[\mathcal{A}]_{\leq i}$. *The dual cone to* $\mathsf{K}_j(g) = \{f \in \mathbb{R}[\mathcal{A}]_{\leq j} \; : \; gf \text{ is } \mathsf{X}\text{-SAGE}\}$ *is*

$$\mathsf{K}_j(g)^{\dagger} = \{v \in \mathbb{R}^{\mathcal{A}_j} \; : \; \exists y \in C_{\mathsf{X}}(\mathcal{A}_{i+j})^{\dagger}, \; v = M_j(g)y\}.$$

Proof. A vector $v \in \mathbb{R}^{\mathcal{A}_j}$ belongs to $\mathsf{K}_j(g)^{\dagger}$ if and only if

$$0 = \inf\{L_v(f) \; : \; gf \text{ is } \mathsf{X}\text{-SAGE}\}.$$

We apply duality to this problem. Since $f \in \mathbb{R}[\mathcal{A}]_{\leq j}$ and $g \in \mathbb{R}[\mathcal{A}]_{\leq i}$, we have that $gf \in \mathbb{R}[\mathcal{A}]_{\leq i+j}$. The constraint that gf is X-SAGE can therefore be associated to a dual variable $y \in C_{\mathsf{X}}(\mathcal{A}_{i+j})^{\dagger}$. We form the Lagrangian $\mathcal{L}(f, y) := L_v(f) - L_y(gf)$ for the above optimization problem (where $f \in \mathbb{R}[\mathcal{A}]_{\leq j}$ is an unconstrained primal variable). Membership of $v \in \mathsf{K}_j(g)^{\dagger}$ is therefore equivalent to the existence of $y \in C_{\mathsf{X}}(\mathcal{A}_{i+j})^{\dagger}$ where $\nabla_f \mathcal{L}(f, y) = \mathbf{0}$. To express this condition explicitly we expand

$$\mathcal{L}(f, y) = \sum_{\alpha \in \mathcal{A}_j} f_\alpha (v_\alpha - L_y(e^\alpha g)),$$

so that $\nabla_f \mathcal{L}(f, y) = \mathbf{0}$ reduces to "$v_\alpha = L_y(e^\alpha g)$ for all $\alpha \in \mathcal{A}_j$". Next, use $L_y(e^\alpha g) = \sum_{\beta \in \mathcal{A}_{i+j}} y_\beta (e^\alpha g)_\beta$. Therefore taking $((e^\alpha g)_\beta \; : \; \beta \in \mathcal{A}_{i+j})$ as row α of $M_j(g)$ for each $\alpha \in \mathcal{A}_j$, we have $\nabla_f \mathcal{L}(f, y) = \mathbf{0}$ if and only if $v = M_j(g)y$. $\square$

Proposition 6.6.4. *Fix* $\ell \in \mathbb{N}_+$, *and for* $j \leq \ell$ *define* $w_j = (\sum_{\alpha \in \mathcal{A}} e^\alpha)^{\ell-j}$. *The dual cone to* $C_{\mathsf{X}}^{\ell}(\mathcal{A})$ *is*

$$C_{\mathsf{X}}^{\ell}(\mathcal{A})^{\dagger} = \left\{v \in \mathbb{R}^{\mathcal{A}_\ell} \; : \; \forall j \leq \ell, \exists y_j \in C_{\mathsf{X}}(\mathcal{A}_\ell)^{\dagger} \text{ where } v = M_j\left(w_j\right) y_j\right\}.$$

Proof. Using the notation from Lemma 6.6.3, we have $C_{\mathsf{X}}^{\ell}(\mathcal{A}) = \sum_{j=1}^{\ell} \mathsf{K}_\ell(w_j)$. Since Minkowski sum and intersection are dual operations we get $C_{\mathsf{X}}^{\ell}(\mathcal{A})^{\dagger} = \cap_{j=1}^{\ell} \mathsf{K}_\ell(w_j)^{\dagger}$. The claim follows by substituting the expressions for $\mathsf{K}_\ell(w_j)^{\dagger}$ obtained from Lemma 6.6.3. $\square$

Chapter 7

CONDITIONAL SAGE FOR POLYNOMIALS

7.1 Introduction

Our first substantive contributions in this thesis were in Chapter 3, where we studied SAGE signomials and introduced the idea of globally nonnegative SAGE polynomials. We have paid little attention to polynomials since then, preferring instead to pursue the deeper questions surrounding the *conditional SAGE signomials* from Chapter 4. This pursuit has been fruitful. We have, through Chapters 5 and 6, come a long way in understanding nonnegative signomials as a fundamental class of functions.

Still, polynomials reign supreme in mathematical modeling. It is not enough to leave the task of jumping from signomials to polynomials as an exercise for the reader. This chapter presents a proper notion of "conditional SAGE polynomials" that we introduced in [60]. We start with an unassuming definition –

a polynomial $x \mapsto \sum_{\alpha \in \mathcal{A}} c_\alpha x^\alpha$ is called X-AGE if it is nonnegative on X and at most one term $c_\beta x^\beta$ attains a negative value on X

– and we proceed by taking sums. Recalling our results from Chapter 3, it is clear that taking $X = \mathbb{R}^n$ recovers our original notion of SAGE polynomials. It is also clear that when $X = \mathbb{R}^n_+$, cones of coefficients for X-AGE polynomials are the same as those of ordinary AGE signomials. What of other cases? Well, that can be tricky. Here it is not convexity of X that determines when an X-SAGE polynomial cone is tractable, but rather convexity of an appropriate logarithmic transform of X.

The organization of this chapter is similar to that of Chapter 4. The remainder of this section presents notation and provides background material. Definitions, representations, and other basic theorems for the conditional SAGE polynomial cones are given in Section 7.2. Section 7.3 addresses solution recovery from dual SAGE relaxations, and Section 7.4 provides a worked example with a special focus on solution recovery. Section 7.5 uses a range of SAGE-based techniques (including ordinary SAGE polynomials) for an example polynomial optimization problem on $\mathbb{R}^n_+$. Further numerical experiments with these SAGE polynomials can be found in Section 8.5.

7.1.1 Notation and definitions

We refer to polynomials by $\text{Pol}(\mathcal{A}, \boldsymbol{c})$ and signomials by $\text{Sig}(\mathcal{A}, \boldsymbol{c})$. The cone of X-AGE signomials supported on $\mathcal{A}$ with free term $\boldsymbol{\beta} \in \mathcal{A}$ is denoted $C_{\text{AGE}}(\mathcal{A}, \text{X}, \boldsymbol{\beta})$ and the resulting cone of X-SAGE signomials is $C_{\text{SAGE}}(\mathcal{A}, \text{X})$. We drop the annotation "X" from those cones to mean $\text{X} = \mathbb{R}^n$.

We use two named machines for our examples, as in Chapter 4. Machine **W** is an HP Z820 workstation, with two 8-core 2.6GHz Intel Xeon E5-2670 processors and 256GB 1600MHz DDR3 RAM. Machine **L** is a 2013 MacBook Pro, with a dual-core 2.4GHz Intel Core i5 processor and 8GB 1600MHz DDR3 RAM.

7.1.2 Background

SAGE signomials can be used to certify global polynomial nonnegativity (see [59] or Chapter 3). For a finite set $\mathcal{A} \subset \mathbb{R}^n$ and a vector $\boldsymbol{c}$ in $\mathbb{R}^{\mathcal{A}}$, we define the set of *signomial representative coefficient vectors* as

$$\text{SR}(\mathcal{A}, \boldsymbol{c}) = \{\hat{\boldsymbol{c}} \in \mathbb{R}^{\mathcal{A}} : \hat{c}_{\alpha} = c_{\alpha} \text{ whenever } \boldsymbol{\alpha} \text{ is in } 2\mathbb{N}^n, \text{ and}$$
$$\hat{c}_{\alpha} \leq -|c_{\alpha}| \text{ whenever } \boldsymbol{\alpha} \text{ is not in } 2\mathbb{N}^n\}.$$

Using a termwise argument, if $\hat{\boldsymbol{c}}$ belongs to $\text{SR}(\mathcal{A}, \boldsymbol{c})$ and $\text{Sig}(\mathcal{A}, \hat{\boldsymbol{c}})$ is nonnegative on $\mathbb{R}^n$, then $\text{Pol}(\mathcal{A}, \boldsymbol{c})$ must likewise be nonnegative on $\mathbb{R}^n$. We define the cone of coefficients for SAGE polynomials as

$$C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A}) \doteq \{\boldsymbol{c} : \text{SR}(\mathcal{A}, \boldsymbol{c}) \cap C_{\text{SAGE}}(\mathcal{A}) \text{ is nonempty }\}. \tag{7.1}$$

Alternatively, one may define an *AGE polynomial* as a nonnegative polynomial $\text{Pol}(\mathcal{A}, \boldsymbol{c})$ where at most one term $c_{\alpha}\boldsymbol{x}^{\alpha}$ attains a negative value as $\boldsymbol{x}$ varies over $\mathbb{R}^n$. Taking sums of such functions will also recover the cone in (7.1).

The theory of ordinary SAGE certificates has connections to a long-running history of similar nonnegativity certificates. The earliest developments here are the *agiforms* introduced by Reznick in 1989 [14]. More recently, Pébay, Rojas, and Thompson studied maximization of circuit functions [93], Pantea, Koeppl, and Craciun introduced *monomial dominating posynomials* [17],[1] and Iliman and de Wolff proposed *sums of nonnegative circuit polynomials* (SONC) [96]. When polynomial SAGE certificates were introduced, it was shown that a polynomial admits a SAGE decomposition if and only if it admits a SONC decomposition [59, Corollary 21]; this led to the first polynomial-time algorithm for optimizing over cones of SONC

---

[1]See also August, Koeppl, and Craciun [29].

polynomials [59, Theorem 16].[2] Most recently, Katthän, Naumann, and Theobald proposed a class of SAGE-like functions which mix polynomials and generalized polynomials [130]; the techniques presented in this chapter apply to such functions with straightforward changes.

Lastly we mention *sums of squares* (SOS) polynomials. A polynomial f is said to be SOS if it can be written in the form $f = \sum_{i=1}^{m} f_i^2$ for appropriate polynomials f_i. In the context of polynomial optimization, one usually parameterizes the SOS cone by a number of variables n and a maximum degree $2d$; this cone can be represented as $\{p : p(x) = L_d^n(x)^\top M L_d^n(x), M \succeq 0\}$ where $L_d^n : \mathbb{R}^n \to \mathbb{R}^{\binom{n+d}{d}}$ is the map from a vector x to the vector of all monomials of degree at-most-d evaluated at x. The connection between SOS-representability and semidefinite programming was first observed by Shor [1], and was subsequently developed by Parrilo [2] and Lasserre [3].

7.2 The conditional SAGE polynomial cones

We call $f = \mathrm{Pol}(\mathcal{A}, c)$ an X-*AGE polynomial* if it is nonnegative over X, and $f(x)$ contains at most one term $c_\beta x^\beta$ which is negative for some x in X. The βth X-AGE polynomial cone is given by

$$C_{\mathsf{AGE}}^{\mathsf{POLY}}(\mathcal{A}, \beta, \mathsf{X}) = \big\{ c \in \mathbb{R}^{\mathcal{A}} : \mathrm{Pol}(\mathcal{A}, c)(x) \geq 0 \text{ for all } x \text{ in } \mathsf{X},$$
$$c_\alpha \geq 0 \text{ if } \alpha \neq \beta \text{ and } x^\alpha > 0 \text{ for some } x \text{ in } \mathsf{X},$$
$$c_\alpha \leq 0 \text{ if } \alpha \neq \beta \text{ and } x^\alpha < 0 \text{ for some } x \text{ in } \mathsf{X} \big\}.$$

Naturally, $f = \mathrm{Pol}(\mathcal{A}, c)$ is an X-*SAGE polynomial* if c belongs to

$$C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X}) \doteq \sum_{\beta \in \mathcal{A}} C_{\mathsf{AGE}}^{\mathsf{POLY}}(\mathcal{A}, \beta, \mathsf{X}).$$

Let us work through some consequences of the definition.

For starters, if x^α takes on positive and negative values as x varies over X, then for every $\beta \in \mathcal{A} \setminus \alpha$ and every $c \in C_{\mathsf{AGE}}^{\mathsf{POLY}}(\mathcal{A}, \beta, \mathsf{X})$, we must have $c_\alpha = 0$. Note that in order for x^α to take on both positive and negative values, α cannot be even ($\alpha \notin 2\mathbb{N}^n$). If X contains an open ball around the origin, then x^α takes on both positive and negative values *if and only if* α is not even. Thus, the definition of X-AGE polynomials agrees with the definition of ordinary AGE polynomials, as proposed in Section 3.5.

---

[2]See [59, Section 5] for discussion on this topic and related results by Wang [100].

Another important case is when X is a subset of the nonnegative orthant. This point is addressed in some detail later in this section; as a preliminary remark, we note that by considering the connection between polynomials and signomials, one can easily see that if $\mathsf{X} \subset \mathbb{R}^n_{++}$ then $C^{\mathsf{POLY}}_{\mathsf{AGE}}(\mathcal{A}, \beta, \mathsf{X}) = C_{\mathsf{AGE}}(\mathcal{A}, \beta, \log \mathsf{X})$.

Many of theorems for signomials from Chapter 4 apply directly to X-SAGE polynomials. For example, it is easy to show the polynomial analog to Proposition 4.2.7: if X is bounded, then $f = \mathrm{Pol}(\mathcal{A}, c)$ with $\mathbf{0} \in \mathcal{A}$ has

$$f^{\mathsf{SAGE}}_{\mathsf{X}} = \sup\{\, \gamma \,:\, \gamma \text{ in } \mathbb{R}, \; c - \gamma \delta_0 \text{ in } C^{\mathsf{POLY}}_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{X})\} > -\infty.$$

Corollary 4.2.2 likewise extends to polynomials. Other than substituting AGE signomial cones with AGE polynomial cones, the only difference is that $\mathcal{N}$ becomes $\mathcal{N} = \{\alpha \in \mathcal{A} : c_\alpha x^\alpha < 0 \text{ for some } x \text{ in } \mathsf{X}\}$.

Now we turn to representation of SAGE polynomial cones. By applying a simple continuity argument one can show that if $\mathsf{X} = \mathrm{cl\,int}\,\mathsf{X} \subset \mathbb{R}^n_+$ then $C^{\mathsf{POLY}}_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{X}) = C_{\mathsf{SAGE}}(\mathcal{A}, \log \mathsf{X})$. This claim is strengthened slightly and made more explicit through the following theorem.

Theorem 7.2.1. *Suppose* $\mathsf{X} = \mathrm{cl}\{x \,:\, \mathbf{0} < x, H(x) \leq \mathbf{1}\}$ *for a continuous map* $H : \mathbb{R}^n \to \mathbb{R}^r$. *Then for* $\mathsf{Y} = \{y \,:\, H(\exp y) \leq \mathbf{1}\}$, *we have* $C^{\mathsf{POLY}}_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{X}) = C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})$.

The proof of Theorem 7.2.1 is straightforward, and hence omitted. A more sophisticated result concerns when X possesses a certain sign-symmetry.

Theorem 7.2.2. *Suppose* $\mathsf{X} = \mathrm{cl}\{x \,:\, \mathbf{0} < |x|, H(|x|) \leq \mathbf{1}\}$ *for a continuous map* $H : \mathbb{R}^n \to \mathbb{R}^r$. *Then for* $\mathsf{Y} = \{y \,:\, H(\exp y) \leq \mathbf{1}\}$, *we have*

$$C^{\mathsf{POLY}}_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{X}) = \{c \,:\, \mathrm{SR}(\mathcal{A}, c) \cap C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y}) \text{ is nonempty }\}. \qquad (7.2)$$

By combining Theorem 4.2.4 with Theorems 7.2.1 and 7.2.2, we know that there exist a range of sets X for which optimization over X-SAGE polynomials is tractable. There remains the potentially nontrivial task of formulating a problem so that one of these theorems provides an efficient representation of $C^{\mathsf{POLY}}_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{X})$; important examples of when this is possible include constraints such as

$$-a \leq x_j \leq a, \qquad \|x\|_p \leq a, \qquad |x^\alpha| \geq a, \quad \text{and} \quad x_j^2 = a$$

where $a > 0$ is a fixed constant.

Theorem 7.2.2. Suppose $c \in C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A}, \mathsf{X})$ is given by a sum $c = \sum_{\beta \in \mathcal{A}} c^{(\beta)}$, where $c^{(\beta)}$ belongs to the βth X-AGE polynomial cone over exponents $\mathcal{A}$. Define $\{\tilde{c}^{(\beta)}\}_{\beta \in \mathcal{A}}$ as follows

$$\tilde{c}_\alpha^{(\beta)} = \begin{cases} -|c_\alpha^{(\beta)}| & \text{if } \beta \text{ is not even, and } \alpha = \beta \\ c_\alpha^{(\beta)} & \text{if otherwise} \end{cases}.$$

By the invariance of X under reflection about hyperplanes $\{x : x_j = 0\}$, and continuity of polynomials, we have that

$$0 \leq \inf\{\text{Pol}(\mathcal{A}, c^{(\beta)})(x) : x \text{ in } \mathsf{X}\} = \inf\{\text{Pol}(\mathcal{A}, \tilde{c}^{(\beta)})(x) : x \text{ in } \mathsf{X} \cap \mathbb{R}_+^n\}$$
$$= \inf\{\text{Sig}(\mathcal{A}, \tilde{c}^{(\beta)})(y) : y \text{ in } \mathsf{Y}\}.$$

The signomials $\text{Sig}(\mathcal{A}, \tilde{c}^{(\beta)})$ are thus nonnegative over $\mathsf{Y} = \{y : H(\exp y) \leq 1\}$, and posses at most one negative coefficient. This implies that $\tilde{c} \doteq \sum_{\beta \in \mathcal{A}} \tilde{c}^{(\beta)}$ belongs to $C_{\text{SAGE}}(\mathcal{A}, \mathsf{Y})$. One may verify that $\tilde{c}$ also satisfies $\tilde{c} \in \text{SR}(\mathcal{A}, c)$, and so we conclude that the right-hand-side of Equation (7.2) contains $C_{\text{SAGE}}^{\text{POLY}}(\mathcal{A}, \mathsf{X})$.

Now we address the reverse inclusion. Let c be such that $\text{SR}(\mathcal{A}, c) \cap C_{\text{SAGE}}(\mathcal{A}, \mathsf{Y})$ is nonempty. One may verify that basic properties of $C_{\text{SAGE}}(\mathcal{A}, \mathsf{Y})$ and $\text{SR}(\mathcal{A}, c)$ ensure that if the intersection is nonempty, it contains an element $\tilde{c}$ satisfying $|c| = |\tilde{c}|$. Henceforth fix $\tilde{c}$ satisfying these conditions. Next we appeal to a relaxed form of Corollary 4.2.2. Setting $\mathcal{N} = \{\beta \in \mathcal{A} : \tilde{c}_\beta \leq 0\}$, there exist vectors $\tilde{c}^{(\beta)}$ satisfying

$$\tilde{c} = \sum_{\beta \in \mathcal{N}} \tilde{c}^{(\beta)}, \quad \tilde{c}^{(\beta)} \in C_{\text{AGE}}(\mathcal{A}, \beta, \mathsf{Y}), \quad \text{and} \quad \tilde{c}_\alpha^{(\beta)} = 0 \text{ for all } \beta \neq \alpha \text{ in } \mathcal{N}.$$

Note the definition of $\text{SR}(\mathcal{A}, c)$ ensures that $\mathcal{N} = \{\beta : \beta \notin 2\mathbb{N}^n, \text{ or } c_\beta \leq 0\}$. Thus we define $c^{(\beta)}$ by

$$c_\alpha^{(\beta)} = \begin{cases} (\text{sgn } c_\alpha)|\tilde{c}_\alpha| & \text{if } \beta \text{ is not even, and } \alpha = \beta \\ \tilde{c}_\alpha^{(\beta)} & \text{if otherwise} \end{cases}$$

so that $c = \sum_{\beta \in \mathcal{N}} c^{(\beta)}$, and each $c^{(\beta)}$ has the necessary sign pattern for membership in the βth AGE cone with respect to $\mathcal{A}, \mathsf{X}$. Finally, note that

$$\inf\{\text{Pol}(\mathcal{A}, c^{(\beta)})(x) : x \text{ in } \mathsf{X}\} = \inf\{\text{Sig}(\mathcal{A}, \tilde{c}^{(\beta)})(y) : y \text{ in } \mathsf{Y}\} \geq 0.$$

to complete the proof. $\qquad\square$

7.3 Solution recovery and sparse moment problems

Just like in the signomial case from Chapter 4, dual SAGE relaxations can be used to recover optimal and near-optimal solutions to polynomial optimization problems. We state a particularly elementary relaxation here. Let $f = \mathrm{Pol}(\mathcal{A}, \boldsymbol{c})$ be the minimization objective and X be some subset of $\mathbb{R}^n$. Additionally, consider a polynomial map G with coordinate functions $g_i = \mathrm{Pol}(\mathcal{A}, \boldsymbol{g}_i)$, and assemble the coefficient vectors $\boldsymbol{g}_i$ into the rows of a matrix $\boldsymbol{G}$. In this notation,

$$\inf\{\langle \boldsymbol{c}, \boldsymbol{v} \rangle \, : \, \boldsymbol{v} \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger}, \, \langle \boldsymbol{\delta_0}, \boldsymbol{v} \rangle = 1, \, \boldsymbol{G}\boldsymbol{v} \geq \boldsymbol{0}\} \tag{7.3}$$

is a convex relaxation of

$$(f, G)_{\mathsf{X}}^{\star} = \inf\{f(\boldsymbol{x}) \, : \, \boldsymbol{x} \in \mathsf{X}, \, G(\boldsymbol{x}) \geq \boldsymbol{0}\}. \tag{7.4}$$

We need a representation for $C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger}$ in order for (7.3) to be useful. We can obtain such representations as corollaries from Theorems 7.2.1 and 7.2.2.

Corollary 7.3.1. *Fix* $\mathsf{Y} = \{\boldsymbol{y} : H(\exp \boldsymbol{y}) \leq \boldsymbol{1}\}$ *for a continuous* $H : \mathbb{R}^n \to \mathbb{R}^r$.

- *If* $\mathsf{X} = \mathrm{cl}\{\boldsymbol{x} \, : \, \boldsymbol{0} < \boldsymbol{x}, \, H(\boldsymbol{x}) \leq \boldsymbol{1}\}$, *then* $C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger} = C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})^{\dagger}$.

- *If* $\mathsf{X} = \mathrm{cl}\{\boldsymbol{x} \, : \, \boldsymbol{0} < |\boldsymbol{x}|, \, H(|\boldsymbol{x}|) \leq \boldsymbol{1}\}$, *then*

$$C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger} = \{\boldsymbol{v} : \textit{there exists } \hat{\boldsymbol{v}} \textit{ in } C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})^{\dagger} \textit{ with}$$
$$|\boldsymbol{v}| \leq \hat{\boldsymbol{v}}, \textit{ and } v_{\alpha} = \hat{v}_{\alpha} \textit{ when } \boldsymbol{\alpha} \in 2\mathbb{N}^n\}.$$

Corollary 7.3.1 holds regardless of whether or not Y is convex. However, to take advantage of it, we need Y to be a tractable convex set. We therefore assume Y is convex for the remainder of the section.

Solution recovery for polynomial optimization is more difficult than for signomial optimization, because monomials possess both signs and magnitudes. We propose a two-phase approach for this problem, where different techniques are used to recover variable magnitudes and variable signs. The main ideas for each phase are described in Sections 7.3.1 and 7.3.2, while the formal algorithms are given in the appendix. The recovered signs and magnitudes are then combined in an elementary way, as given by the following algorithm.

Algorithm 2 solution recovery for dual SAGE polynomial relaxations.

Input: An objective polynomial f and a polynomial map G supported on exponents $\mathcal{A}$. Vectors $v \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger}$ and $\hat{v} \in C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})^{\dagger}$. Tolerances $\epsilon_{\mathrm{con}}, \epsilon_0 > 0$.

 1: **procedure** POLYSOLUTIONRECOVERY($f, G, \mathcal{A}, v, \hat{v}, \epsilon_{\mathrm{con}}, \epsilon_0$)
 2:     M $\leftarrow$ VariableMagnitudes($\mathcal{A}$, v, $\hat{v}$, ϵ_0).   # Algorithm 3
 3:     S $\leftarrow$ {**1**}
 4:     **if** X is not a subset of $\mathbb{R}_+^n$ **then**
 5:         S.union(VariableSigns($\mathcal{A}$, v))   # Algorithm 4
 6:     solutions $\leftarrow$ [].
 7:     **for** x_{mag} in M and s in S **do**
 8:         $x \leftarrow x_{\mathrm{mag}} * s$ # denotes elementwise multiplication
 9:         **if** $G(x) \geq -\epsilon_{\mathrm{con}} \cdot \mathbf{1}$ **then**
10:             solutions.append(x)
11:     solutions.sort(f, increasing).
12:     **return** solutions.

If v is optimal for an appropriate SAGE relaxation and $v = (x^{\alpha} : \alpha \in \mathcal{A})$ for an elementwise nonzero x in X, then Algorithm 2 will return an optimal solution to problem (7.4).

7.3.1 Recovering variable magnitudes

Given a $v \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger}$, we want to find an $x \in \mathsf{X}$ satisfying $(x^{\alpha})_{\alpha \in \mathcal{A}} = |v|$.

Regardless of whether X is sign-symmetric or contained in the nonnegative orthant, the variable $v \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})$ is associated with an auxiliary variable $\hat{v} \in C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})$, and the variable $\hat{v}$ is associated with additional auxiliary variables z_{β} as part of the standard representation for dual Y-AGE signomial cones. As we discussed in Section 4.4, the vectors $y_{\beta} = z_{\beta}/\hat{v}_{\beta}$ belong to Y, and so the vectors $x_{\beta} = \exp y_{\beta}$ must belong to X. These vectors x_{β} are not only feasible with respect to X, but also satisfy $(x^{\alpha})_{\alpha \in \mathcal{A}} = \hat{v}$ under the binding-constraint and normalization conditions alluded to in Section 4.4. Since $\hat{v} = |v|$ always holds at least for $\mathsf{X} \subset \mathbb{R}_+^n$, the vectors $x_{\beta} = \exp(z_{\beta}/\hat{v}_{\beta})$ are reasonable candidates for variable magnitudes.

It is possible that $|v| \neq \hat{v}$ when X is sign-symmetric. This is particularly likely when v is subject to additional linear constraints, such as $Gv \geq 0$. Therefore when X is sign-symmetric, it is worth considering variable magnitudes which supplement the ones described above. We propose that one picks a threshold $\epsilon_0 > 0$, computes

$$y \in \mathrm{argmin}\{ \textstyle\sum_{\alpha:v_{\alpha} \neq 0}(\langle \alpha, y \rangle - \log |v_{\alpha}|)^2 : y \text{ in } \mathsf{Y}, \tag{7.5}$$

$$\langle \alpha, y \rangle \leq \log(\epsilon_0) \text{ for all } v_{\alpha} = 0\},$$

and exponentiates $x = \exp y$. The role of ϵ_0 is to ensure x satisfies $|x|^\alpha \le \epsilon_0$ whenever $v_\alpha = 0$. Values of ϵ_0 below machine precision are reasonable here.

A formal statement of our method for magnitude recovery (Algorithm 3) can be found in the appendix.

7.3.2 Recovering variable signs

In this subsection we consider $\mathcal{A}$ as a tall $m \times n$ matrix. Row i of $\mathcal{A}$ is α_i and the jth entry of α_i is α_{ij}.

Let $\mathsf{M}^{-1}(v)$ denote the set of $x \in \mathbb{R}^n$ satisfying $v = (x^{\alpha_1}, \ldots, x^{\alpha_m})$. Henceforth, fix v and assume $\mathsf{M}^{-1}(v)$ is nonempty. Here we describe how to find vectors s in $\{+1, 0, -1\}^n$ so that at least one $x \in \mathsf{M}^{-1}(v)$ satisfies $x_i > 0$ when $s_i = +1$, $x_i = 0$ when $s_i = 0$, and $x_i < 0$ when $s_i = -1$. Once we describe this process, we relax the problem slightly so $s_i = +1$ allows $x_i = 0$.

First we address when s_i should equal zero. Let $\mathsf{U} = \{i \in [m] : v_i \ne 0\}$. Consider how if some $x \in \mathsf{M}^{-1}(v)$ has $x_j = 0$, then we must have $\alpha_{ij} = 0$ for all i in U (else $x^{\alpha_i} = v_i \ne 0$ would fail). Thus when $\alpha_{ij} = 0$ for all i in U, we set $s_j = 0$ without loss of generality. Now let $\mathsf{W} = \{j \in [n] : \alpha_{ij} > 0 \text{ for some } i \text{ in } \mathsf{U}\}$; these are indices for which s_j is not yet decided. Consider the vector $(v < 0) \in \{0, 1\}^n$ with values $(v < 0)_i = 1$ if $v_i < 0$, and zero if otherwise. Let $\mathcal{A}[\mathsf{U}, :]$ be the submatrix of $\mathcal{A}$ formed by rows $\{\alpha_i\}_{i \in \mathsf{U}}$, and similarly index $(v < 0)$. Finally, solve

$$\mathcal{A}[\mathsf{U}, :]z \equiv (v < 0)[\mathsf{U}] \mod 2 \quad \text{and} \quad z_j = 0 \text{ for all } j \text{ in } [n] \setminus \mathsf{W} \quad (7.6)$$

for z in $\{0, 1\}^n$. The remaining $(s_j)_{j \in \mathsf{W}}$ are $s_j = -1$ if $z_j = 1$ and $s_j = 1$ otherwise.

An individual solution to (7.6) can be computed efficiently by Gaussian elimination over the finite field $\mathbb{F}_2$. Our formal algorithm for solution recovery provides the option to recover all solutions to (7.6), using additional techniques from finite-field linear algebra (c.f. [168]). See the appendix for details.

7.4 An example with sign-symmetric constraints

This section's example is to minimize a function appearing in the formulation of the cyclic n-roots problem. The general cyclic n-roots problem is a challenging benchmark problem in computer algebra [169]. Our problem is to minimize

$$f(x) = -64 \sum_{i=1}^{7} \prod_{j \in [7] \setminus \{i\}} x_j \quad \text{(Ex3)}$$

over the box $\mathsf{X} = [-1/2, 1/2]^7$. To our knowledge, this problem was first used as an optimization benchmark in the work by Ray and Nataraj, on computing the extrema of polynomials over boxes [170]. One may verify that $f_{\mathsf{X}}^\star = -7$, and that this objective value is attained at $\boldsymbol{x}^{(1)} = \boldsymbol{1}/2$ and $\boldsymbol{x}^{(2)} = -\boldsymbol{1}/2$. Despite this problem's simplicity, it requires nontrivial computational effort with SOS methods. The lowest relaxation order that allows GloptiPoly3 [121] to compute $f_{\mathsf{X}}^\star = -7$ results in a semidefinite program that takes MOSEK 90 seconds to solve with Machine **W**.

SAGE relaxations automatically exploit the structure in this problem. Since the seven functions $f_i(\boldsymbol{x}) = 1 - 64 \prod_{j \neq i} x_j$ are X-AGE and sum to $f + 7$, we have that $-7 \leq f_{\mathsf{X}}^{\mathsf{SAGE}} \leq f_{\mathsf{X}}^\star$. To address the dual SAGE relaxation and solution recovery, we introduce the matrix $\mathcal{A} = (\alpha_{ij} : (i, j) \in [8] \times [7])$, with final row $\alpha_8 = \boldsymbol{0}$, $\alpha_{ii} = 0$ for $i \leq 7$, and $\alpha_{ij} = 1$ for the remaining entries. Next we write $\mathsf{X} = \{\boldsymbol{x} : \boldsymbol{x}^2 \leq 1/4\}$, and for $\mathsf{Y} = \{\boldsymbol{y} : \exp(2\boldsymbol{y}) \leq \boldsymbol{1}/4\}$ numerically solve

$$f_{\mathsf{X}}^{\mathsf{SAGE}} = \inf\{-64 \cdot \boldsymbol{1}^\top \boldsymbol{v}_{1:7} : -\hat{\boldsymbol{v}} \leq \boldsymbol{v} \leq \hat{\boldsymbol{v}},$$

$$\hat{\boldsymbol{v}} \text{ in } C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})^\dagger, \ v_8 = \hat{v}_8 = 1\} = -7.$$

MOSEK solves this problem in 0.01 seconds with Machine **W**.

We recover candidate magnitudes by using the eight Y-AGE cones associated with the auxiliary variable $\hat{\boldsymbol{v}} \in C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})^\dagger$. To machine precision, each of these AGE cones yields the same candidate magnitude $|\boldsymbol{x}| = \boldsymbol{1}/2$. The optimal moment vector $\boldsymbol{v} = \boldsymbol{1}/64$ is elementwise positive, and so sign-pattern recovery is a matter of finding all solutions to the system $\mathcal{A}\boldsymbol{z} \equiv \boldsymbol{0} \mod 2$. There are exactly two solutions to this system: $\boldsymbol{z}^{(1)} = \boldsymbol{0}$, and $\boldsymbol{z}^{(2)} = \boldsymbol{1}$. The first of these gives rise to signs $\boldsymbol{s}^{(1)} = \boldsymbol{1}$, and the second of these results in $\boldsymbol{s}^{(2)} = -\boldsymbol{1}$. By combining these candidate signs with candidate magnitudes, we obtain candidate solutions $\{\boldsymbol{1}/2, -\boldsymbol{1}/2\}$; since these solutions are feasible and obtain objective values matching the SAGE bound, we conclude that both candidate solutions are minimizers of f over X.

7.5 An example with nonnegative decision variables

Our next problem appears in work on *bounded degree sums of squares* (BSOS) and *sparse bounded degree sums of squares* (Sparse-BSOS) methods for polynomial optimization [171, 172]. The latter paper reports BSOS and Sparse-BSOS compute $(f, g)_{\mathbb{R}^6}^\star = -0.41288$ in 44.5 and 82.1 seconds respectively, when using SDPT3-4.0 on a machine with a 4-core 2.6GHz Core i7 processor and 16GB RAM.

$$\inf_{\boldsymbol{x}\in\mathbb{R}^6}\ f(\boldsymbol{x}) \doteq x_1^6 - x_2^6 + x_3^6 - x_4^6 + x_5^6 - x_6^6 + x_1 - x_2 \quad \text{subject to} \tag{7.7}$$

$$g_1(\boldsymbol{x}) \doteq 2x_1^6 + 3x_2^2 + 2x_1x_2 + 2x_3^6 + 3x_4^2 + 2x_3x_4 + 2x_5^6 + 3x_6^2 + 2x_5x_6 \geq 0$$

$$g_2(\boldsymbol{x}) \doteq 2x_1^2 + 5x_2^2 + 3x_1x_2 + 2x_3^2 + 5x_4^2 + 3x_3x_4 + 2x_5^2 + 5x_6^2 + 3x_5x_6 \geq 0$$

$$g_3(\boldsymbol{x}) \doteq 3x_1^2 + 2x_2^2 - 4x_1x_2 + 3x_3^2 + 2x_4^2 - 4x_3x_4 + 3x_5^2 + 2x_6^2 - 4x_5x_6 \geq 0$$

$$g_4(\boldsymbol{x}) \doteq x_1^2 + 6x_2^2 - 4x_1x_2 + x_3^2 + 6x_4^2 - 4x_3x_4 + x_5^2 + 6x_6^2 - 4x_5x_6 \geq 0$$

$$g_5(\boldsymbol{x}) \doteq x_1^2 + 4x_2^6 - 3x_1x_2 + x_3^2 + 4x_4^6 - 3x_3x_4 + x_5^2 + 4x_6^6 - 3x_5x_6 \geq 0$$

$$g_{6:10}(\boldsymbol{x}) \doteq 1 - g_{1:5}(\boldsymbol{x}) \geq \boldsymbol{0}$$

$$g_{11:16}(\boldsymbol{x}) \doteq \boldsymbol{x} \geq \boldsymbol{0}$$

We note that Problem (7.7) is very sparse; it includes only 22 of the $\binom{12}{6} = 924$ distinct monomials that could appear in a degree 6 polynomial optimization problem in 6 variables. This problem is also a good example for conditional SAGE polynomials, because it allows for several choices in partial dualization.

We approach this problem with a particular level of a hierarchy of SAGE relaxations for polynomial optimization. Rather than state the hierarchy in full generality (see Subsection 8.5.1), we say that a pair $(\tilde{G}, \mathsf{X})$ induces a bound $(f, \tilde{G})_{\mathsf{X}}^{\text{SAGE}}$ equal to

$$\sup\ \gamma \ \text{s.t.}\ f - \gamma - \sum_{\tilde{g}\in\tilde{G}} s_{\tilde{g}} \cdot \tilde{g} \ \text{is an X-SAGE polynomial}$$

$$s_{\tilde{g}} \ \text{are X-SAGE polynomials over exponents } \mathcal{A} \cup (2\mathcal{A}).$$

In the formulation above, $\mathcal{A} \subset \mathbb{N}^n$ is the smallest set for which every polynomial in $\{f,\ \boldsymbol{x} \mapsto 1\} \cup \tilde{G}$ is in the span of the monomial basis $\{\boldsymbol{x} \mapsto \boldsymbol{x}^\alpha\}_{\alpha\in\mathcal{A}}$.

The simplest way to approach Problem (7.7) to use no partial dualization at all– simply take $\mathsf{X} = \mathbb{R}^n$. Indeed, it is possible to solve Problem (7.7) with only these ordinary SAGE certificates, however computing $(f, g)_{\mathbb{R}^6}^{\text{SAGE}} = -0.41288$ requires 101 seconds of solver time on Machine **W**.

A preferable alternative is to use partial dualization with $\mathsf{X} = \mathbb{R}_+^6$. With this choice of X it is natural to drop now trivially-satisfied constraints from g, and work with $\hat{g} = g_{3:10}$. This allows us to compute $(f, \hat{g})_{\mathsf{X}}^{\text{SAGE}} = -0.41288$ in 3.04 seconds of solver time on Machine **W**, and 4.4 seconds of solver time on Machine **L**. Significantly, the SAGE relaxation solve time on Machine **L** is an order of magnitude smaller than the BSOS solve time reported in [172].

The most aggressive choice for partial dualization is $\mathsf{X} = \{x : x \geq 0, g_{6:7}(x) \geq 0\}$. With this choice of X one may use $\hat{g} = (g_{3:5}, g_{8:10})$, or $\hat{g} = g_{3:10}$; in the first case Machine **W** computes $(f, \hat{g})^{\mathrm{SAGE}}_{\mathsf{X}} = -0.47121$ in 3.3 seconds, and in the second case Machine **W** computes $(f, \hat{g})^{\mathrm{SAGE}}_{\mathsf{X}} = -0.41288$ in 5.67 seconds.

C h a p t e r   8

THE SAGEOPT PYTHON PACKAGE

8.1   Introduction

This chapter describes the `sageopt`[1] python package and some experiments we conducted with this package in [60].

These experiments are separated into groups of signomial optimization and polynomial optimization (Sections 8.4 and 8.5). In both cases we make use of certain hierarchies of SAGE-based convex relaxations. These hierarchies were designed in a heuristic way by borrowing ideas from [13] and the signomial hierarchy does not incorporate our advances in Chapter 6. Still, even the heuristically designed hierarchy was successful in solving polynomial optimization problems with practical relevance in electrical engineering [63, 64].

All experiments were conducted on Machine **W** using MOSEK 9.0.70(beta). Subsections 8.4.1 and 8.5.1 only state the SAGE relaxations in primal form. However, these experiments were conducted by symbolically constructing primal and dual problems, and solving them separately from one another. In order to communicate the quality of these numeric solutions, we generally report "SAGE bounds" to the farthest decimal point where the primal and dual objectives agree.

Remark 8.1.1. This chapter has several named "Example" problems that are written out in full. These examples are numbered starting with "Example 5." The first four examples have actually appeared in Chapters 4 and 7. We did not re-index these examples because we wanted to make it easy to cross-reference with [60].

8.2   About `sageopt`

`Sageopt` is designed so the mathematics in this thesis translates into python code with as little modification as possible. We implement everything "in-house," including symbolic `Signomial` and `Polynomial` objects, and an algebraic modeling language for convex optimization. The convex optimization modeling system supports explicit primal and dual SAGE cone constraints. The SAGE constraints are part of the public API and have data structures to manage presolve and dimension-reduction techniques.

---

[1]Also acceptably styled as `SAGEopt`.

It should be straightforward for a control theorist to perform Lyapunov analysis using only the symbolic `Signomial` and `Polynomial` objects and `sageopt`'s convex optimization modeling system. However, the situation is more complicated when we get to using SAGE for nonconvex signomial and polynomial optimization. `Sageopt` therefore includes several pre-defined hierarchies of convex relaxations for such problems. The hierarchies are given in primal and dual forms, and solution recovery algorithms are implemented for the dual forms. At present, the particular hierarchies are given in Section 8.4 (for signomials) and Subsection 8.5.2 (for polynomials).

Here is an example showing how easy it is to use `sageopt`.

```python
import sageopt as so
import numpy as np

# Define problem data: min f(x) subject to g(x) >= 0
t = so.standard_sig_monomials(3)   # t = exp(x)
f = 0.5 * t[0] / t[1] - t[0] - 5 / t[1]
gts = [100 -   t[1] / t[2] - t[1] - 0.05 * t[0] * t[2],
       150 - t[0], t[0] - 70,
        30 - t[1],  t[1] - 1,
        21 - t[2],  t[2] - 0.5]
X = so.infer_domain(f, gts, [])

# Construct, solve, and post-process a SAGE relaxation
prob = so.sig_relaxation(f, X)
simple_bound = prob.solve()
x_opt = so.sig_solrec(prob)[0]
t_opt = np.exp(x_star)

# Certify that the recovered point was optimal
strong_bound = so.sig_relaxation(f, X, ell=3).solve()
```

Listing 8.1: Using `sageopt` to solve the problem in (4.8)

Some unconstrained optimization features of `sageopt` have counterparts in the POEM python package [99, 109]. GloptiPoly3 [121] and SOSTOOLS [173] are the SOS counterparts to `sageopt`.

8.2.1   Integration with GPKit

`Sageopt`'s main limitation is that it requires users to work with vectorized models. You can only have a single decision variable x in some subset of $\mathbb{R}^n$. This limitation is not very restrictive given the expressive data structures available in

python, but it could still be a hurdle that prevents an a practicing engineer from using sageopt. To overcome this limitation, we collaborated with Berk Ozturk do develop a basic interface with the GPKit modeling platform. GPKit is a popular tool for optimization-driven engineering design [174] and is even used at Virgin Hyperloop to create large-scale signomial programming models [56].

Here is a toy example that showcases sageopt's integration with GPKit.

```python
import sageopt as so
import numpy as np
from sageopt.interop.gpkit import gpkit_model_to_sageopt_model
from gpkit import Variable, Model, SignomialsEnabled
from gpkit.constraints.sigeq import SingleSignomialEquality
#
# Build a toy GPKit model (for illustrative purposes)
#
x = Variable('x')
y = Variable('y')
with SignomialsEnabled():
    constrs = [0.2 <= x,
               x <= 0.95,
               SingleSignomialEquality(x + y, 1)]
gpkm = Model(x*y, constraints)
#
#    Recover data for the sageopt model
#
som = gpkit_model_to_sageopt_model(gpkm)  # a dictionary
eqs, gts = som['sp_eqs'], som['gp_gts']
f = som['f']
X = so.infer_domain(f, gts, [])
prob = so.sig_constrained_relaxation(f, gts, eqs, X, p=1)
#
#    Solve and recover solution
#
prob.solve(solver='ECOS', verbose=False)
soln = so.sig_solrec(prob)[0]
geo_soln = np.exp(soln)
vkmap = som['vkmap']
x_val = geo_soln[vkmap[x.key]]
y_val = geo_soln[vkmap[y.key]]
```

Listing 8.2: Using sageopt with GPKit

The master branch of sageopt currently only supports GPKit versions below 1.0.

We have a branch on `sageopt`'s GitHub repository that is ready for when `GPKit` 1.0 is released to the public. Future plans for integration with these two packages include a proper data structure for signomial programs within `sageopt` (rather than requiring lists of `Signomial` objects). Such a data structure has been prepared as part of our work in Chapter 6, but we have yet to move that work into the public `sageopt` repository.

8.2.2 `Coniclifts` in `sageopt` 0.5

`Sageopt` does not solve SAGE relaxations on its own; it relies on third-party convex optimization solvers, such as ECOS [107, 115] or MOSEK [119]. These solvers require input in very specific standard-forms. `Coniclifts` provides abstractions that allow us to state SAGE relaxations in high-level syntax, and manage interactions with these low-level solvers.

SAGE constraints in `coniclifts`

`Coniclifts` includes direct implementations of primal and dual SAGE cones, which have virtually identical constructors and public attributes. These classes also share a common data structure called an `ExpCoverHelper`, to ensure that certain presolve procedures are applied in a symmetric way for both primal and dual SAGE constraints.

One aspect of presolve is easiest to state with primal SAGE constraints. Suppose c is an affine expression of some decision variables γ, λ. If certain components of c are constant with respect to γ, λ, then we can take advantage of the fixed sign of those components to reduce the number of AGE cones used in a constraint "$c \in C_X(\mathcal{A})$." This presolve procedure is justified by Theorem 3.3.1 and its corollaries.

`Coniclifts` also maintains several options for controlling exactly how a primal or dual SAGE constraint is compiled into a low-level standard form. Some of these questions are mundane, such as when to introduce slack variables and when to relax an equality constraint to an inequality without loss of generality. A more significant question is whether to eliminate the equality constraints of the form $\mathcal{A}^\dagger v = 0$ in primal SAGE constraints as in our proof of Theorem 3.5.1.

`Coniclifts`' design

`Coniclifts` is built around a few core ideas, including

- transparency in the problem compilation process,

- ease-of-extension for experts in convex optimization,

- no dependence on a C or C++ backend,

- full compatibility with NumPy.

In order to achieve full compatibility with NumPy, `coniclifts` takes an elementwise approach to symbolic expressions. Specifically, we begin with a few simple abstractions for scalar-valued symbolic expressions, and wraps those abstractions in a custom subclass of NumPy's "ndarray." The `coniclifts` abstractions for scalar-valued symbolic expressions are as follows:

- A `ScalarExpression` class represents scalar-valued affine functions of certain irreducible primitives. `ScalarExpression`s are operator-overloaded to support +, −, and ∗. This allows ndarrays of `ScalarExpression`s to fall back on many functions which are implemented for numeric ndarrays.

- An abstract `ScalarAtom` class specifies the behavior of the irreducible primitives in `ScalarExpression`s. The `ScalarAtom` class immediately specializes into `ScalarVariable`s (far and away the most important `ScalarAtom`) and `NonlinearScalarAtom`s. `NonlinearScalarAtom`s are implemented on a case-by-case basis, but include such things as the exponential function and the vector 2-norm.

To our (*very limited!*) knowledge, `coniclifts` is the only algebraic modeling system for optimization built directly on top of NumPy.

8.3 Principles for optimization via nonnegativity certificates

We are motivated by a desire to solve optimization problems

$$(f, G)_{\mathsf{X}}^{\star} = \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \text{ in } \mathsf{X} \subset \mathbb{R}^n, \ G(\boldsymbol{x}) \geq \boldsymbol{0}\} \tag{8.1}$$

where f is a function from $\mathbb{R}^n$ to $\mathbb{R}$ and G is a vector-valued map on $\mathbb{R}^n$. Our primary goal is to produce lower bounds $(f, G)_{\mathsf{X}}^{\mathrm{lb}} \leq (f, G)_{\mathsf{X}}^{\star}$. In the event that $(f, G)_{\mathsf{X}}^{\mathrm{lb}} = (f, G)_{\mathsf{X}}^{\star}$, we are also interested in recovering optimal solutions to (8.1).

Remark 8.3.1. Equality constraints can nominally be handled by two-sided inequalities. Actual implementations of the techniques described in this chapter (for example, as provided in `sageopt`) treat equality constraints in more efficient ways.

We begin by reviewing the Lagrange dual relaxation of the above problem, both in minimax form and as a nonnegativity problem. From there we review standard techniques for strengthening nonnegativity-based relaxations of problems such as (8.1); this includes the use of redundant constraints, nonconstant Lagrange multipliers, and strengthening nonnegativity certificates via modulation. We then introduce partial dualization. Until we discuss partial dualization, the set X appearing in Problem 8.1 shall be the whole of $\mathbb{R}^n$.

8.3.1 Dual problems in nonconvex optimization

The simplest way to lower bound $(f, G)^\star_{\mathbb{R}^n}$ is via the Lagrange dual. For each coordinate function g of G, we introduce a dual variable $\lambda_g \in \mathbb{R}_+$ and consider the Lagrangian $L(\boldsymbol{x}, \boldsymbol{\lambda}) = f(\boldsymbol{x}) - \langle \boldsymbol{\lambda}, G(\boldsymbol{x}) \rangle$. The *Lagrange dual problem* is to compute

$$(f, G)^{\mathsf{L}}_{\mathbb{R}^n} = \sup_{\boldsymbol{\lambda} \geq \boldsymbol{0}} \inf_{\boldsymbol{x} \in \mathbb{R}^n} L(\boldsymbol{x}, \boldsymbol{\lambda}).$$

By the minimax inequality, we can be certain that $(f, G)^{\mathsf{L}}_{\mathbb{R}^n} \leq (f, G)^\star_{\mathbb{R}^n}$.

There are many situations when the Lagrange dual problem is intractable. For signomial and polynomial optimization, one usually needs to compute yet another lower bound $(f, G)^{\mathsf{d}}_{\mathbb{R}^n} \leq (f, G)^{\mathsf{L}}_{\mathbb{R}^n}$. We start by introducing a parameterized function $\psi(\gamma, \boldsymbol{\lambda})$ which takes values $\psi(\gamma, \boldsymbol{\lambda})(\boldsymbol{x}) = L(\boldsymbol{x}, \boldsymbol{\lambda}) - \gamma$. One reformulates the dual as

$$(f, G)^{\mathsf{L}}_{\mathbb{R}^n} = \sup\{\gamma \ : \ \boldsymbol{\lambda} \geq \boldsymbol{0}, \ \gamma \text{ in } \mathbb{R}, \ \psi(\gamma, \boldsymbol{\lambda})(\boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{x} \text{ in } \mathbb{R}^n\},$$

and the constraint that "$\psi(\gamma, \boldsymbol{\lambda})$ defines a nonnegative function" is then tightened to "$\psi(\gamma, \boldsymbol{\lambda})$ satisfies a particular sufficient condition for nonnegativity." The expectation is that the sufficient condition can be expressed by tractable convex constraints on variables γ and $\boldsymbol{\lambda}$. For example, SOS certificates for polynomial nonnegativity can be expressed via linear matrix inequalities, and SAGE certificates for signomial and polynomial nonnegativity can be expressed with the relative entropy function.

8.3.2 Strengthening dual bounds in nonnegativity relaxations

A common method for strengthening dual problems is to introduce redundant constraints to the primal problem, particularly by taking products of existing constraint functions. As an example of this principle in action, consider the toy polynomial optimization problem

$$\inf\{ -x^2 \ : \ -1 \leq x \leq 1 \} = -1.$$

One may verify that $(f, G)_{\mathbb{R}}^L = -\infty$, but by adding the single redundant constraint $(1 - x)(1 + x) \geq 0$, we can certify a dual bound $-1 \leq (f, G)_{\mathbb{R}}^{\star}$.

A more subtle method is to reconsider what is meant by "dual variables." For the Lagrange dual problem we use scalars $\lambda_g \in \mathbb{R}_+$, however it is just as valid to have λ_g be a *function*, provided it is nonnegative over $\mathbb{R}^n$. Such a method is well-suited to our nonnegativity-based relaxations of the dual problem. The following toy signomial program illustrates the utility of this approach

$$\inf\{-\exp(2x) \ : \ 1 \leq \exp(x) \leq 2\} = -4.$$

Again the Lagrange dual problem returns a bound of $-\infty$, but by considering $\lambda_i(x) = \eta_i \exp(x)$ with $\eta_i \geq 0$, the resulting dual bound is $-4 \leq (f, G)_{\mathbb{R}}^{\star}$.

A third method for strengthening dual bounds only becomes relevant when working with strict inner-approximations of nonnegativity cones. For two functions w, f with w positive definite, it is clear that f is nonnegative if and only if the product $w \cdot f$ is nonnegative. The *method of modulation* is to choose a generic positive-definite function w so that if f fails a particular test for nonnegativity (say, being SOS, or being SAGE), there is still a chance that the product $w \cdot f$ passes a test for nonnegativity. Indeed, modulation is a crucial tool for computing successive bounds for unconstrained problems

$$f_{\mathbb{R}^n}^{\star} := \inf\{f(\boldsymbol{x}) : \boldsymbol{x} \text{ in } \mathbb{R}^n\} = \sup\{\gamma : f(\boldsymbol{x}) - \gamma \geq 0 \text{ for all } \boldsymbol{x} \text{ in } \mathbb{R}^n\}.$$

Suppose for example that f is a signomial over exponents $\mathcal{A}$; then for $w = \text{Sig}(\mathcal{A}, \mathbf{1})$ we can compute a non-decreasing sequence of lower bounds

$$f_{\mathbb{R}^n}^{(\ell)} = \sup\{\gamma : \gamma \text{ in } \mathbb{R}, \ w^\ell(f - \gamma) \text{ is SAGE}\} \leq f_{\mathbb{R}^n}^{\star}.$$

For suitable conditions on $\mathcal{A}$ (c.f. [13, Theorem 4.1]), we have $f_{\mathbb{R}^n}^{(\ell)} \to f_{\mathbb{R}^n}^{\star}$ as ℓ tends to infinity. From an implementation perspective, the constraint that "$\psi(\gamma) := w^\ell(f - \gamma)$ is SAGE" is tractable because the coefficient vector of $\psi(\gamma)$ is an affine function of γ.

Modulation can similarly be applied to constrained optimization. Suppose that $L(\boldsymbol{x}, \lambda)$ is the Lagrangian for Problem 8.1, and refer to the function $\boldsymbol{x} \mapsto L(\boldsymbol{x}, \lambda)$ as $L(\lambda)$. Then rather than requiring that "$L(\lambda) - \gamma$ is SAGE," one can require that "$\psi(\gamma, \lambda) := w^\ell(L(\lambda) - \gamma)$ is SAGE." This increases the size of the feasible set for variables γ and λ, and remains tractable due to the affine dependence of $\psi(\gamma, \lambda)$ on γ and λ. Such modulation leads to a non-decreasing sequence of bounds which converge to $(f, G)_{\mathbb{R}^n}^L$ under suitable conditions.

Remark 8.3.2. Our framing of the method of modulation is meant to keep things simple. However, in the real algebraic geometry literature, it would be more appropriate to say that we are looking for certificates of nonnegativity involving *rational functions*. For example, rather than asking that f be SOS, we can ask that f is a sum of squares of rational functions.

8.3.3 Partial dualization

Partial dualization is a technique for strengthening dual bounds, which is at least as strong as any choice of redundant constraints or nonconstant Lagrange multipliers. Considering Problem (8.1) now with $X \subsetneq \mathbb{R}^n$, the natural generalization of the Lagrange dual is

$$(f, G)_X^d := \sup\{ \gamma : \lambda \geq 0, \ \gamma \text{ in } \mathbb{R}, \ L(x, \lambda) - \gamma \geq 0 \text{ for all } x \text{ in } X\}. \qquad (8.2)$$

In the important case when X is compact, we are guaranteed to have $(f, G)_X^d > -\infty$, a property which is in stark contrast to the Lagrange dual. We call (8.2) a *partial dual* if $X = \{x : g_i(x) \geq 0 \text{ for all } i \text{ in } I\}$ was constructed from some subset $I \subset [k]$ of the constraint functions. Note that in the extreme case with $X = \{x : G(x) \geq 0\}$, we have $(f, 0)_X^d = (f, G)_{\mathbb{R}^n}^\star$. This is to say, partial dualization provides a mechanism to completely eliminate duality gaps.

We now provide a simple example that combines partial dualization and nonnegativity certificates. Suppose we want to minimize a univariate polynomial f over an interval $[a, b]$, subject to a polynomial inequality constraint $g(x) \geq 0$. In this case we may form a Lagrangian $L(x, \lambda) = f(x) - \lambda g(x)$ with $\lambda \geq 0$, and find the largest constant γ so that $x \mapsto L(x, \lambda) - \gamma$ is nonnegative over $x \in [a, b]$. A result by Powers and Reznick states that a degree-d polynomial "p" is nonnegative over an interval $[a, b]$ if and only if it can be written as $p(x) = s(x)^2 + h_{[a,b]}(x)t(x)^2$, where $h_{[a,b]}(x) = (b - x)(x - a)$, and s, t are polynomials of degree at most d and $d - 1$ respectively [175]. Therefore the partial dual $(f, g)_{[a,b]}^d$ can be framed as an SOS relaxation

$$(f, g)_{[a,b]}^d = \sup\{\gamma : f - \lambda g - \gamma = \tilde{s} + \tilde{t} \cdot h_{[a,b]}, \ \lambda \geq 0,$$
$$\tilde{s} \in \text{SOS}(2d), \ \tilde{t} \in \text{SOS}(2(d - 1))\}.$$

Where we have used "SOS($2d$)" to denote the cone of SOS polynomials in one variable that are of degree at most $2d$. This example is particularly nice, as all nonnegative univariate polynomials are actually SOS.

Our last concept is how partial dualization manifests in the dual of the dual, also known as the *moment relaxation.* In describing this concept we use the notation from Chapter 7. Fix $\mathcal{A} \subset \mathbb{R}^n$ with $\mathbf{0} \in \mathcal{A}$ and a set $\mathsf{X} \subset \mathbb{R}^n$. Consider an objective $f = \text{Pol}(\mathcal{A}, \boldsymbol{c})$ and constraint functions $G = \{\text{Pol}(\mathcal{A}, \boldsymbol{g}_i)\}_{i \in [k]}$, and assemble the coefficient vectors $\boldsymbol{g}_i$ to form the rows of a matrix $\boldsymbol{G} \in \mathbb{R}^{k \times \mathcal{A}}$. The partial dual $(f, G)_{\mathsf{X}}^{\text{d}}$ can be written as the convex cone program

$$(f, G)_{\mathsf{X}}^{\text{d}} = \sup\{\gamma \ : \ \boldsymbol{\lambda} \geq \mathbf{0}, \ \gamma \in \mathbb{R}, \ \boldsymbol{c} - \gamma \boldsymbol{\delta_0} - \boldsymbol{G}^\dagger \boldsymbol{\lambda} \in C_{\text{NNP}}(\mathcal{A}, \mathsf{X})\},$$

and we can apply conic duality to obtain

$$(f, G)_{\mathsf{X}}^{\text{d}} = \inf\{\langle \boldsymbol{c}\, \boldsymbol{v} \rangle \ : \ \langle \boldsymbol{v}, \boldsymbol{\delta_0} \rangle = 1, \ \boldsymbol{Gv} \geq \mathbf{0}, \ \boldsymbol{v} \in C_{\text{NNP}}(\mathcal{A}, \mathsf{X})^\dagger\}.$$

The set $C_{\text{NNP}}(\mathcal{A}, \mathsf{X})^\dagger$ is the closed cone generated by all vector-valued expectations $\mathbb{E}_{\boldsymbol{x} \sim F}[(\boldsymbol{x}^{\boldsymbol{\alpha}} \ : \ \boldsymbol{\alpha} \in \mathcal{A})]$, where F is a probability measure conditioned on $\boldsymbol{x} \in \mathsf{X}$. As we saw in Chapters 4 and 7, moment relaxations can be used in solution recovery schemes to certify $(f, G)_{\mathsf{X}}^{\text{d}} = (f, G)_{\mathsf{X}}^\star$.

8.4 Signomial optimization

The examples in this section were drawn from the PhD thesis of James Yan [86], a popular benchmarking paper by Rijckaert and Martens [162], and the more contemporary works [83, 84]. This section is organized chronologically with respect to these sources. Many of the problems considered here can be found elsewhere in the literature; see Shen et al. [77, 80, 82], Wang and Liang [78], and Qu et al. [81].

SAGE recovers best-known solutions for all but six of the twenty-nine problems considered here. For every one of these six problematic examples, numerical issues resulted in solver failures for level-(p, q, ℓ) relaxations whenever $p > 0$; the results for these six problems should not be taken as definitive. For the twenty-three problems where SAGE recovered best-known solutions, there are two important trends we can observe. First, our solution recovery algorithms are more likely to succeed with a conditional SAGE relaxation than with an ordinary SAGE relaxation, even when the ordinary SAGE relaxation is tight. Second, solution refinement by the COBYLA local solver can help tremendously in the presence of suboptimal strictly-feasible initial solutions (Example 8), and in the presence of both large and small constraint violations (Examples 9 and 6 respectively). The initial condition from a SAGE relaxation in local refinement is important; the underlying COBYLA solver can and will return suboptimal solutions if initialized poorly.

8.4.1 Two reference hierarchies for signomial optimization

Here we describe a particular set of choices for two SAGE-based hierarchies for signomial programming. When we say "$\{f\} \cup G$ are signomials over exponents $\mathcal{A}$," we mean that $\{x \mapsto \exp\langle\alpha, x\rangle\}_{\alpha \in \mathcal{A}}$ is the smallest monomial basis spanning all linear combinations of f, $g \in G$, and the function $x \mapsto 1$.

First we describe a SAGE-based hierarchy that does not use the minimax inequality, i.e., a hierarchy applicable when all constraints can be moved into X. Formally, for a signomial f over exponents $\mathcal{A}$, a set $\mathsf{X} \subset \mathbb{R}^n$, and an integer $\ell \geq 0$, the *level-ℓ SAGE relaxation* for $f_{\mathsf{X}}^{\star}$ is

$$f_{\mathsf{X}}^{(\ell)} := \sup\{\gamma \,:\, \mathrm{Sig}(\mathcal{A}, \mathbf{1})^{\ell}(f - \gamma) \text{ is X-SAGE}\}. \tag{8.3}$$

The special case with $\ell = 0$ is sometimes denoted "$f_{\mathsf{X}}^{\mathsf{SAGE}}$."

Now we consider functional constraints; let $\{f\} \cup G$ be a set of signomials over exponents $\mathcal{A}$. SAGE relaxations for the problem of computing $(f, G)_{\mathsf{X}}^{\star}$ are indexed by three integer parameters: p, q, and ℓ. Starting from $p \geq 0$ and $q \geq 1$, define $\mathcal{A}[p]$ as the matrix of exponent vectors for $\mathrm{Sig}(\mathcal{A}, \mathbf{1})^p$, and define $G[q]$ as the set of all products of at-most-q elements of g. The SAGE relaxation for $(f, G)_{\mathsf{X}}^{\star}$ at level (p, q, ℓ) is then

$$
\begin{aligned}
(f, G)_{\mathsf{X}}^{(p,q,\ell)} = \sup \quad & \gamma \text{ s.t. } s_h \text{ are signomials over exponents } \mathcal{A}[p]\\
& L := f - \gamma - \sum_{h \in G[q]} s_h \cdot h\\
& \mathrm{Sig}(\mathcal{A}, \mathbf{1})^{\ell} L \text{ is an X-SAGE signomial}\\
& s_h \text{ are X-SAGE signomials.}
\end{aligned}
\tag{8.4}
$$

The decision variables in (8.4) are $\gamma \in \mathbb{R}$ and the coefficient vectors of $\{s_h\}_{h \in G[q]}$. The most basic level of this hierarchy is $(p, q, \ell) = (0, 1, 0)$. This corresponds to using scalar Lagrange multipliers ($s_h \geq 0$), the original constraints ($G[0] = G$), and modulating the Lagrangian by the signomial that is identically equal to 1. Note that when $p > 0$, the Lagrange multipliers s_h need only be nonnegative on X, rather than over the whole of $\mathbb{R}^n$.

Remark 8.4.1. It is often useful to apply a local solver to the output of Algorithm 1. The term "Algorithm 1L" henceforth refers to the use of Algorithm 1, followed by solution refinement with Powell's COBYLA solver [165].[2] Our later experiments

---

[2]A FORTRAN implementation is accessible through `SciPy`'s `optimize` submodule. We use this implementation with `RHOBEG=1`, `RHOEND=` 10^{-7}, and `MAXFUN=` 10^5.

show that equality constraints can stymie solution recovery even when SAGE computes $(f, G)_{\mathsf{X}}^{\star}$. We therefore suggest that one eliminate equality constraints through substitution of monomials $\exp(x_i)$, when possible. Alternatively, one can allow large violations of any problematic constraints in Algorithm 1, and pass the returned values as near-feasible points to a local solver, as in Algorithm 1L.

8.4.2 Problems from the PhD thesis of James Yan

We attempted to solve nine example problems appearing James Yan's 1976 PhD thesis *Signomial programs with equality constraints : numerical solution and applications* [86]. This section reproduces two of the six problems which we solved to global optimality via SAGE certificates. Yan's "Problem B" (page 88) and "Problem C" (page 89) serve as our Examples 5 and 6 respectively.

First we consider Example 5

$$\inf_{\boldsymbol{x} \in \mathbb{R}^4} f(\boldsymbol{x}) := 2 - \exp(x_1 + x_2 + x_3) \tag{Ex5}$$

$$\text{s.t. } g_1(\boldsymbol{x}) := 4 - \exp x_3 - 15\exp(x_2 + x_3) - 1.5\exp(x_3 + x_4) \geq 0$$

$$g_{2:5}(\boldsymbol{x}) := (1, 1, 1, 2) - \exp \boldsymbol{x} \geq \boldsymbol{0}$$

$$g_{6:9}(\boldsymbol{x}) := \exp \boldsymbol{x} - (1, 1, 1, 1)/10 \geq \boldsymbol{0}$$

$$\phi_1(\boldsymbol{x}) := \exp x_1 + 2\exp x_2 + 2\exp x_3 - \exp x_4 = 0.$$

It is possible to quickly compute $(f, g, \phi)_{\mathbb{R}^4}^{\star} = 1.\overline{925}$ with both conditional and ordinary SAGE certificates, although conditional SAGE certificates exhibit better performance for solution recovery. Specifically, $(f, g, \phi)_{\mathbb{R}^4}^{(1,1,0)} = 1.92592592$ can be computed in 0.12 seconds, but no solution can be recovered from Algorithm 1 unless ϵ is set to an unacceptably large value of 0.1. Instead we set $\mathsf{X} = \{\boldsymbol{x} : g(\boldsymbol{x}) \geq \boldsymbol{0}\}$, compute $(f, g, \phi)_{\mathsf{X}}^{(1,1,0)} = 1.92592593$ in 0.18 seconds, and by running Algorithm 1 recover $\boldsymbol{x}^{\star}$ satisfying $g(\boldsymbol{x}^{\star}) > 1\text{E-}11$, $|\phi(\boldsymbol{x}^{\star})| < 1\text{E-}8$, and $f(\boldsymbol{x}^{\star}) = 1.92592593$.

Remark 8.4.2. The formulation of Example 5 given here differs from that in [60]. Specifically, we have corrected a typo in [60] that had the constant "1.5" in the defining expression for g_1 incorrectly given as "15." This difference is a typographical error; all SAGE bounds and runtimes here are the same as in [60].

Now we turn to Example 6

$$\inf_{\mathbf{y}\in\mathbb{R}^3_{++}} \; y_1^{0.6}y_2 + y_2y_3^{-0.5} + 15.98y_1 + 9.0824y_2^2 - 60.72625y_3 \qquad\text{(Ex6)}$$

$$\text{s.t. } y_2^{-2}y_3 - y_1y_2^{-2} - 0.48 \geq 0$$

$$y_1^{0.5}y_3^2 - y_1^{0.25}y_3 - y_2^2 - 5.75 \geq 0$$

$$(1000, 1000, 1000) \geq \mathbf{y} \geq (0.1, 0.1, 0.1)$$

$$y_1^2 + 4y_2^2 + 2y_3^2 - 58 = 0$$

$$y_1y_2^{-1}y_3^{2.5} + y_2y_3 - y_2^2 - 16.55 = 0.$$

With $\mathsf{X} = \{\mathbf{x} \in \mathbb{R}^3 : g(\mathbf{x}) \geq \mathbf{0}\}$, we can compute $(f, g_{1:2}, \phi)_{\mathsf{X}}^{(0,1,0)} = -320.722913$ in 0.04 seconds. By running Algorithm 1 with $\epsilon_{\text{ineq}} =$1E-8 and $\epsilon_{\text{eq}} =$1E-6, we recover $\mathbf{x}$ with objective $f(\mathbf{x}) = -320.722913$ and that is feasible up to tolerance 8E-7. We then pass this solution to COBYLA with `RHOEND=`1E-10, and subsequently recover $\mathbf{x}^\star$ with the same objective, but a constraint violation of only 5E-13.

The remaining problems which we solved to optimality were "Problem A" on page 60, "Problem A" on page 88, "Problem D" on page 89, and the problem in equation environment "(6.15)" on page 106. The last of these was introduced in Section 4.6. The problems which we did not solve to optimality were "Problem B" on page 61, the problem in equation environment "(6.29)" on page 113, and the problem in equation environment "(6.36)" on page 120. In each of these unsolved cases, we encountered solver-failures for level-(p, q, ℓ) relaxations whenever $p > 0$. Therefore the bounds computed for each of these problems were essentially limited to those of Lagrange dual problems, with modest partial dualization.

8.4.3 Problems from the benchmarking paper of Rijckaert and Martens

We consider problems 9 through 18 of the popular signomial-geometric programming benchmark paper by Rijckaert and Martens [162]. Of these ten problems, seven met with at least moderate success, in that SAGE relaxations produced meaningful lower bounds on a problem's optimal value, and also facilitated recovery of best-known solutions to these problems. SAGE certificates allow us to certify global optimality for four of these seven problems. Problem statistics and a summary of SAGE performance is given in Table 8.1.

We reproduce Rijckaert and Martens' problems 10 and 15 as our Examples 7 and 8

respectively; both problems are written in exponential-form

$$\inf_{\boldsymbol{x} \in \mathbb{R}^3} f(\boldsymbol{x}) := 0.5 \exp(x_1 - x_2) - \exp x_1 - 5 \exp(-x_2) \tag{Ex7}$$

$$\text{s.t. } g_1(\boldsymbol{x}) := 100 - \exp(x_2 - x_3) - \exp x_1 - 0.05 \exp(x_1 + x_3) \geq 0$$

$$g_{2:4}(\boldsymbol{x}) := (100, 100, 100) - \exp \boldsymbol{x} \geq \boldsymbol{0}$$

$$g_{5:7}(\boldsymbol{x}) := \exp \boldsymbol{x} - (1, 1, 1) \geq \boldsymbol{0}.$$

The bound constraints appearing in Example 7 are not included in [162], however f is unbounded below if we omit them. The solution proposed in [162] has $\exp \boldsymbol{x} = (88.310, 7.454, 1.311)$, and objective value $f(\boldsymbol{x}) = -83.06$. The actual optimal solution has value -83.25, and this can be certified by running Algorithm 1 on a dual solution for $f_{\mathsf{X}}^{(3)} = -83.2510$, where $\mathsf{X} = \{\boldsymbol{x} : g(\boldsymbol{x}) \geq \boldsymbol{0}\}$. Solving the necessary SAGE relaxation takes 0.1 seconds on Machine **W**.

Now we have Example 8

$$\inf_{\boldsymbol{x} \in \mathbb{R}^{10}} f(\boldsymbol{x}) := 0.05 \exp x_1 + 0.05 \exp x_2 + 0.05 \exp x_3 + \exp x_9 \tag{Ex8}$$

$$\text{s.t. } g_1(\boldsymbol{x}) := 1 + 0.5 \exp(x_1 + x_4 - x_7) - \exp(x_{10} - x_7) \geq 0$$

$$g_2(\boldsymbol{x}) := 1 + 0.5 \exp(x_2 + x_5 - x_8) - \exp(x_7 - x_8) \geq 0$$

$$g_3(\boldsymbol{x}) := 1 + 0.5 \exp(x_3 + x_6 - x_9) - \exp(x_8 - x_9) \geq 0$$

$$g_4(\boldsymbol{x}) := 1 - 0.25 \exp(-x_{10}) - 0.5 \exp(x_9 - x_{10}) \geq 0$$

$$g_5(\boldsymbol{x}) := 1 - 0.79681 \exp(x_4 - x_7) \geq 0$$

$$g_6(\boldsymbol{x}) := 1 - 0.79681 \exp(x_5 - x_8) \geq 0$$

$$g_7(\boldsymbol{x}) := 1 - 0.79681 \exp(x_6 - x_9) \geq 0.$$

A level $(1,1,0)$ ordinary SAGE relaxation for Example 8 can be solved in 2.8 seconds on Machine **W**; this returns the bound $(f, g)_{\mathbb{R}^{10}}^{(1,1,0)} = 0.2056534$. When Algorithm 1 is run on the dual solution, it returns a point $\boldsymbol{x}$ satisfying $f(\boldsymbol{x}) \approx 0.38$ and $g(\boldsymbol{x}) \geq 0.053$. However by subsequently running Algorithm 1L, we obtain $\boldsymbol{x}^\star$ satisfying $f(\boldsymbol{x}^\star) = 0.20565341$ and $g_i(\boldsymbol{x}^\star) \geq 1\text{E-8}$ for all i in $[k]$. We thus conclude that the level-$(1, 1, 0)$ SAGE relaxation was tight.

Remark 8.4.3. In [60], the number "11.9643" in the first row of Table 8.2 was incorrectly given as "11.9600." We have two other remarks on this problem. First, by applying a monotonicity analysis, it's easy to show that the optimal objective is in fact 11.9643. That is, Algorithm 1L does return the optimal solution in this case, even though our SAGE bound does not tell us as much. Second, the reported

Num. in [162]	n	k	solution quality	optimal?
9	2	1	same	unknown
10	3	1	improved	yes
11	4	2	same	yes
12	8	4	same	unknown
13	8	6	no solution	no
14	10	7	same	unknown
15	10	7	same	yes
16	10	7	same	yes
17	11	8	no solution	no
18	13	9	no solution	no

Table 8.1: Columns n and k give number of variables and number of inequality constraints for the indicated problem. "Solution quality" is "same" (resp. "improved") if Algorithm 1L returned a feasible solution with objective equal to (resp. less than) that proposed in [162]. Problems 9, 12, and 14 are discussed in Table 8.2. We encountered solver failures for level-$(1, 1, 0)$ relaxations of problems 13, 17, and 18.

	SAGE relaxation		Algorithm 1		Algorithm 1L	
Num. in [162]	(p, q, ℓ)	bound	$f(\boldsymbol{x})$	min $g(\boldsymbol{x})$	$f(\boldsymbol{x})$	min $g(\boldsymbol{x})$
9	(3,3,1)	11.7	12.500	0.00438	11.9643	2.00E-10
12	(0,2,1)	-6.4	-5.7677	0.00034	-6.0482	-5.00E-09
14	(0,4,0)	0.7	2.5798	0.01541	1.14396	-8.00E-09

Table 8.2: Problems for which we did not certify optimality, but nevertheless recovered best-known solutions by using SAGE relaxations. Note that Algorithm 1 returned strictly-feasible solutions in each of these cases. In the next section we present examples where Algorithm 1 does not return feasible solutions, and so solution refinement (i.e. Algorithm 1L) becomes more important.

objective value in the paper by Rijckaert and Martens is 11.91, but if you actually evaluate the objective function at the point provided by Rijckaert and Martens, you get 11.96392 and a constraint violation of 5E-6.

8.4.4   Problems from contemporary sources

Here we describe our attempts at solving six problems from the 2014 article by Hou, Shen, and Chen [83], as well as four problems from the 2014 article by Xu [84]. SAGE relaxations are quite successful in this regard: seven of the ten problems are solved to global optimality (verified SAGE bounds), while best-known (but possibly suboptimal) solutions are obtained for the remaining three problems. Summary results can be found in Tables 8.3 and 8.4.

source	num.	n	k_1	k_2	objective	infeasibility	optimal?
[83]	1	4	10	0	0.7650822	0	yes
-	2	2	5	0	11.964337	0	yes
-	3	3	7	0	-147.66666	0	yes
-	5	5	16	0	10122.493	4.00E-13	unknown
-	7	3	6	0	-10.363636	2.00E-15	yes
-	8	15	37	6	156.21963	4.00E-14	yes
[84]	4	2	1	1	1.3934649	2.00E-10	yes
-	5	6	9	4	-0.3888114	5.00E-17	unknown
-	6	2	4	2	1.1771243	4.00E-12	yes
-	7	6	20	3	10252.790	8.00E-14	unknown

Table 8.3: Columns n, k_1, and k_2 specify the number of variables, inequality constraints, and equality constraints in the indicated problem. The last three columns specify the objective value and constraint violation of a solution obtained by running Algorithm 1L on the output of a dual SAGE relaxation, as well as a note on whether the objective matched a SAGE bound. Problems with "unknown" optimality status are described in Table 8.4.

We explicitly reproduce problem [83]-8 as our Example 9

$$\inf_{\boldsymbol{y} \in \mathbb{R}_{++}^{15}} \sum_{i=1}^{4} y_{i+11}(12.62626 - 1.231059 y_i) \tag{Ex9}$$

s.t. $y_{12} - y_{11} \le 0, \quad y_{11} - y_{12} \le 50, \quad y_{10} - y_4 \le 0$

$y_9 - y_{10} \le 0, \quad y_8 - y_9 \le 0, \quad 2y_7 - y_1 \le 1$

$y_3 - y_4 \le 0, \quad y_2 - y_3 \le 0, \quad y_1 - y_2 \le 0$

$50y_4 + y_{10}y_{15} - 50y_{10} - y_4 y_{15} \le 0$

$50y_{10} + y_4 y_5 + y_9 y_{14} - 50y_9 - y_3 y_{14} - y_8 y_{15} \le 0$

$50y_7 + y_2 y_{13} + y_7 y_{12} - 50y_8 - y_1 y_{12} - y_8 y_{13} \le 0$

$50y_8 + y_1 y_{12} + y_8 y_{13} - 50y_7 - y_2 y_{13} - y_7 y_{12} \le 0$

$50y_8 + 50y_9 + y_3 y_{14} + y_8 y_{13} - y_2 y_{13} - y_9 y_{14} \le 500$

$y_6 y_{11} + y_1 y_{12} + y_7 y_{11} - y_6 y_{12} \le 0$

$100 y_{i+5} + 0.0975 y_i^2 - 3.475 y_i - 9.75 y_i y_{i+5} \le 0$ for all i in [5]

$\boldsymbol{y} \ge (1.000000, 1, 9, 9, 9, 1, 1.000000, 1, 1, 1, 50, 0.0, 1.0, 50, 50)$

$\boldsymbol{y} \le (8.037732, 9, 9, 9, 9, 1, 4.518866, 9, 9, 9, 100, 50, 50, 50, 50).$

Six of the fifteen variables in Example 9 have matching upper and lower bounds – these are the six equality constraints alluded to in Table 8.3. Our formulation differs from [83]-8, in that a constraint "$x_3 x_2 - x_3 \le 0$" in the original problem statement

was replaced by "$y_2 - y_3 \leq 0$" in our problem statement. This change is necessary because the original problem is actually infeasible.

We approach Example 9 by maximizing our use of partial dualization: the set $\mathsf{X} \subset \mathbb{R}^{15}$ includes all bound constraints, all but two of the first nine inequality constraints, as well as the constraint fourth from the end of the problem statement. The equality constraints implied for variables $y_3, y_4, y_5, y_6, y_{14}, y_{15}$ are not included in the Lagrangian. A level-(0,1,0) conditional SAGE relaxation then produces a bound $(f, g)_{\mathsf{X}}^{\star} \geq 156.2196$ in 0.05 seconds. By running Algorithm 1L with $\epsilon_{\text{ineq}} = 100$, we subsequently obtain the solution

$$y_{1:8}^{\star} = (8.037732, 9, 9, 9, 9, 1, 1, 1.15686275)$$
$$y_{9:15}^{\star} = (1.21505203, 1.58987319, 50, 3\text{E-}50, 1, 50, 50).$$

The solution $y^{\star}$ is feasible up to forward-error 3.6E-14, and attains an objective value of 156.219629. Because this objective matches the SAGE bound, we conclude that $y^{\star}$ is optimal up to relative error 2E-7.

source-num.	(p, q, ℓ)	bound	ϵ_{ineq}	ϵ_{eq}	objective	infeasibility
[83]-5	(0,1,0)	9171.00	1.00E-08	0	10122.493	4.00E-13
[84]-5	(2,2,0)	-0.390	1.00E-08	1	-0.3888114	5.00E-17
[84]-7	(0,1,0)	9397.8	1.00E-08	1	10252.790	8.00E-14

Table 8.4: Signomial programs for which we did not certify optimality, but nevertheless recovered best-known solutions by using SAGE relaxations. Columns ϵ_{ineq} and ϵ_{eq} indicate the value of infeasibility tolerances when running Algorithm 1, prior to feeding the output of Algorithm 1 to COBYLA as part of Algorithm 1L. The last two columns list the objective function value and constraint violations for the output of Algorithm 1L. [84]-7 reports a solution $\tilde{x}$ with smaller objective value, however that solution violates an equality constraint $\phi(x) = 0$ with error $|\phi(\tilde{x})| > 0.11$.

8.5 Polynomial optimization

Remark 8.5.1. As with Algorithm 1 in the signomial case, it is useful to apply a simple local solver to the output of Algorithm 2 as a sort of solution refinement. We use the term "Algorithm 2L" in reference to such a method, with COBYLA as the local solver.

8.5.1 Two reference hierarchies for polynomial optimization

If $\mathsf{X} \subset \mathbb{R}_+^n$, then one should use the same hierarchies described in Subsection 8.4.1, where "Sig" is replaced by "Pol" and constraints that a function is "an X-SAGE

signomial" are replaced by constraints that the function is "an X-SAGE polynomial." This section focuses on the more complicated case when X is sign-symmetric.

Our reference hierarchy for functionally constrained polynomial optimization is similar to that used for signomial programming. Let f, $\{g_i\}_{i=1}^{k_1}$, and $\{\phi_i\}_{i=1}^{k_2}$ be polynomials over common exponents $\mathcal{A} \in \mathbb{N}^{m \times n}$, and fix sign-symmetric $\mathsf{X} \subset \mathbb{R}^n$. Define $\hat{\mathcal{A}}$ as the matrix formed by stacking $\mathcal{A}$ on top of $2\mathcal{A}$, and then removing any duplicate rows. The SAGE relaxation for $(f, g, \phi)_{\mathsf{X}}^{\star}$ at level (p, q, ℓ) is then

$$(f, g, \phi)_{\mathsf{X}}^{(p,q,\ell)} = \sup \ \gamma \ \text{s.t.} \ s_h, z_h \text{ are polynomials over exponents } \hat{\mathcal{A}}[p]$$
$$L := f - \gamma - \sum_{h \in g[q]} s_h \cdot h - \sum_{h \in \phi[q]} z_h \cdot h$$
$$\text{Pol}(2\mathcal{A}, \mathbf{1})^{\ell} L \text{ is an X-SAGE polynomial}$$
$$s_h \text{ are X-SAGE polynomials.} \tag{8.5}$$

As before, the decision variables are $\gamma \in \mathbb{R}$, and the coefficient vectors of $\{s_h\}_{h \in g[q]}$, $\{z_h\}_{h \in \phi[q]}$. The main difference between (8.5) and it's signomial equivalent (8.4), is that the Lagrange multipliers are slightly more complex in (8.5). This change was made to improve performance for problems where only a few rows of $\mathcal{A}$ belong to the nonnegative even integer lattice.

Our minimax-free reference hierarchy for polynomial optimization is meaningfully different from the signomial case. We begin by assuming a representation $\mathsf{X} = \text{cl}\{x \ : \ \mathbf{0} < |x|, \ H(|x|) \leq \mathbf{1}\}$, and subsequently defining $\mathsf{Y} = \{y \ : \ H(\exp y) \leq \mathbf{1}\}$. Let A and C be operators on polynomials so that $f = \text{Pol}(A(f), C(f))$ always holds, and let s be the vector in $\mathbb{R}^m$ with $s_i = 1$ when α_i is even, and $s_i = 0$ otherwise. The SAGE relaxation for $f_{\mathsf{X}}^{\star}$ at level (p, q) is

$$f_{\mathsf{X}}^{(p,q)} = \sup \ \gamma \ \text{s.t.} \ \psi := \text{Pol}(\mathcal{A}, s)^p (f - \gamma) \tag{8.6}$$
$$c \in \text{SR}(A(\psi), C(\psi))$$
$$[\text{Sig}(A(\psi), \mathbf{1})]^q \, \text{Sig}(A(\psi), c) \text{ is Y-SAGE}$$

over optimization variables c and γ.

Formulation 8.6 uses two parameters out of desire to mitigate *both* sources of error in the SAGE polynomial cone: that attributable to the use of signomial representatives, and that attributable to the gap between Y-SAGE and Y-nonnegativity. As we show in Subsection 8.5.2, the signomial representative complexity parameter "q" can make the difference in our ability to solve problems when $\mathsf{X} = \mathbb{R}^n$.

8.5.2 Polynomial optimization problems from the literature

Here we review results of the reference hierarchies from Subsection 8.5.1, as applied to twenty-two polynomial optimization problems from the literature. We begin with six unconstrained and eight box-constrained problems (drawn from [176] and [170] respectively). There are two important lessons which we highlight with the box-constrained problems. First, bound constraints should still be included in the Lagrangian, even if they can be completely absorbed into the set "X" in a conditional SAGE relaxation. Second, even if the original problem does not feature many sign-symmetric constraints, it is often easy to infer valid sign-symmetric constraints which can improve performance of a conditional SAGE relaxation. The remaining eight problems discussed in this section have nonconvex objectives, nonconvex inequality constraints, and constraints that the optimization variables are nonnegative [171]. Our experience with such problems is that partial dualization plays a crucial role in solving them efficiently, primarily with the simpler constraints $x \geq 0$.

Table 8.5 describes the unconstrained and box-constrained problems; three such problems are reproduced here, as our Examples 10 through 12.

$$\inf\{f(x) := 4x_1^2 - 2.1x_1^4 + x_1^6/3 + x_1x_2 - 4x_2^2 + 4x_2^4 \; : \; x \text{ in } \mathbb{R}^2\} \qquad \text{(Ex10)}$$

The polynomial f in Example 10 is known as the six-hump camel function; its minimum is $f_{\mathbb{R}^2}^\star \approx -1.0316$. By using polynomial modulators, a level-(3,0) relaxation returns a bound -1.03170 in 0.63 seconds of solver time on Machine **W**. By instead solving a level-(0,2) relaxation (i.e., modulating the signomial representative of $f - \gamma$) we obtain $-1.031630 \leq f_{\mathbb{R}^2}^\star$ in 0.19 seconds. Example 10 shows how the two-parameter hierarchy (8.6) can be of practical importance.

Our next two examples are box-constrained problems from the work of Ray and Nataraj [170]; their problems "Capresse 4" and "Butcher 6" serve as our Examples 11 and 12. A consistent trend for these problems is that even when a feasible set X can be incorporated entirely into an X-SAGE cone, it is still useful to take products of constraints, and solve a relaxation such as (8.5) which includes those constraints in the Lagrangian.

$$\inf_{x \in \mathbb{R}^4} \quad f(x) := -x_1x_3^3 + 4x_2x_3^2x_4 + 4x_1x_3x_4^2 + 2x_2x_4^3 + 4x_1x_3 \qquad \text{(Ex11)}$$
$$+ 4x_3^2 - 10x_2x_4 - 10x_4^2 + 2$$
$$\text{s.t. } g_{1:4}(x) := x + (1, 1, 1, 1)/2 \geq 0$$
$$g_{5:8}(x) := (1, 1, 1, 1)/2 - x \geq 0$$

Letting $\mathsf{X} = \{\boldsymbol{x} \in \mathbb{R}^4 \; : \; -0.5 \leq x_i \leq 0.5\}$, one can compute $(f, g)_{\mathsf{X}}^{(1,2,0)} =$ -3.180096, where the equality is verified by recovering a solution with Algorithm 2. Example 11 is noteworthy because the recovered solution required no local-solver refinement that occurs in Algorithm 2L.

$$\inf_{\boldsymbol{x} \in \mathbb{R}^6} \quad f(\boldsymbol{x}) := x_6 x_2^2 + x_5 x_3^2 - x_1 x_4^2 + x_4^3 + x_4^2 - 1/3 x_1 + 4/3 x_4 \qquad \text{(Ex12)}$$

$$\text{s.t. } g_{1:6}(\boldsymbol{x}) := \boldsymbol{x} + (1, \; 0.1, \; 0.1, \; 1, \; 0.1, \; 0.1) \geq \boldsymbol{0}$$

$$g_{7:12}(\boldsymbol{x}) := (0, \; 0.9, \; 0.5, -0.1, -0.05, -0.03) - \boldsymbol{x} \geq \boldsymbol{0}$$

We can produce a tight bound for Example 12 with ordinary SAGE certificates: a level-(0,3,0) relaxation returns $-1.4392999 \leq (f, g)^{\star}$ in 0.67 seconds. Solution recovery is not so easy. Unless we move to a computationally expensive level-(0,3,1) ordinary SAGE relaxation, Algorithm 2 fails to return a feasible point. Instead, we infer valid inequalities for use in a conditional SAGE relaxation:

$$|x_1| \leq 1, \quad |x_2| \leq 0.9, \quad |x_3| \leq 0.5, \quad \text{and}$$

$$0.1 \leq |x_4| \leq 1, \quad 0.05 \leq |x_5| \leq 0.1, \quad 0.03 \leq |x_6| \leq 0.1.$$

The resulting level (0,3,0) relaxation can be solved in 0.64 seconds. We recover a feasible solution with Algorithm 2, which matches the SAGE bound after refinement by COBYLA. Example 12 reinforces a message from signomial optimization: even if an ordinary SAGE relaxation can produce a tight bound, a conditional SAGE relaxation is likely to fare better with solution recovery. Example 12 also shows how useful sign-symmetric constraints can be inferred from bound constraints which are not sign-symmetric.

Now we turn to problems featuring nonconvex inequality constraints [171]. One of these problems was introduced in Section 7.5 as "Example 4," and all of these problems have a similar structure to that of Example 4. Most importantly, problems featured here include nonnegativity constraints $\boldsymbol{x} \geq \boldsymbol{0}$. The natural SAGE hierarchy solves all of these problems; see Table 8.6.

There are a few subtle distinctions between signomial programs (SPs) with generalized polynomials, and the nonnegative polynomial optimization problems (POPs) considered here. While a polynomial optimization problem over $\boldsymbol{x} \geq \boldsymbol{0}$ may include $x_i = 0$ in the feasible set, generalized-polynomial SPs cannot allow this (since there is the possibility of dividing by zero, or encountering indeterminate forms). Thus solution recovery from SAGE relaxations is nominally more challenging for a true

source	name	n	d	minimum	SAGE solved
[176]	Rosenbrock	variable	4	0	yes
-	6-hump camel	2	6	-1.0316	yes
-	3-hump camel	2	6	0	yes
-	Beale	2	8	0	no
-	Colville	4	4	0	no
-	Goldstein-Price	2	8	3	no
[170]	L.V. 4	4	4	-20.8	yes
-	Cap 4	4	4	-3.117690	yes
-	Hun 5	5	7	-1436.515	no
-	Cyc 5	5	4	-3	yes
-	C.D. 6	6	2	-270397.4	no
-	But 6	6	3	-1.4393	yes
-	Heart 8	8	4	-1.367754	yes
-	Viras 8	8	2	-29	yes

Table 8.5: Results for SAGE on unconstrained and box-constrained polynomial minimization problems. Column "d" indicates the degree of the polynomial to be minimized. The Rosenbrock example allows for different numbers of variables, though results from [59] show SAGE is tight for any number of variables. The Beale, Colville, and Goldstein-Price polynomials proved very difficult for optimization via SAGE certificates.

nonnegative POP, relative to a generalized polynomial SP. Despite this challenge, Algorithm 2L successfully recovers optimal solutions for all of these problems. See Table 8.7 for details.

The other important distinction between generalized-polynomial SPs and nonnegative POPs, is that there exist established SOS based methods for dealing with nonnegative POPs. Thus it is useful to understand the performance of SAGE-based methods in the context of SOS-based methods for polynomial optimization. Although SAGE relaxations took a very long time to solve problems P4_6 and P4_8, the runtimes for problems such as P6_8 are remarkable. The unspecified machine in [171] took over 1600 and 200 seconds to solve P6_8 with BSOS and SOS respectively, while SAGE can solve the same problem in under 4 seconds on a mid-tier laptop from 2013. It seems to the authors that SAGE provides a compelling option for nonnegative polynomial optimization problems, at least for low levels of the reference hierarchy (such as $(1, 1, 0)$, or $(0, q, 0)$ with small q).

name	k	minimum	(p, q, ℓ)	W time (s)	L time (s)
P4_4	8	-0.033538	(1,1,0)	0.47	0.7
P4_6	7	-0.060693	(1,1,1)	289	292
P4_8	7	-0.085813	(2,1,0)	396	460
P6_4	8	-0.576959	(1,1,0)	3.45	4.1
P6_6	8	-0.412878	(1,1,0)	3.04	4.37
P6_8	8	-0.409020	(1,1,0)	3.25	3.83
P8_4	8	-0.436026	(1,1,0)	7.18	7.25
P8_6	8	-0.412878	(1,1,0)	8.67	8.21

Table 8.6: Generic polynomial optimization problems, over the nonnegative orthant [171, 172]. Names "Pn_d" indicate the number of variables n and degree d of the given problem. Column k gives the number of inequality constraints, excluding constraints $x \geq 0$, as well as those which trivially follow from $x \geq 0$. SAGE solved all problems listed here, at the indicated level of the hierarchy, and with the indicated solver runtimes for the primal-form relaxations.

	Algorithm 2		Algorithm 2L	
name	$f(x)$	$\min g(x)$	$f(x)$	$\min g(x)$
P4_4	-0.033386	0.00E-00	-0.033538	0.00E-00
P4_6	-0.057164	4.06E-02	-0.060693	-2.44E-14
P4_8	-0.066671	1.42E-01	-0.085813	-3.46E-14
P6_4	-0.570848	4.04E-08	-0.576959	-1.03E-13
P6_6	-0.412878	5.46E-09	-0.412878	-1.68E-13
P6_8	-0.409018	1.07E-07	-0.409020	-5.82E-14
P8_4	-0.436024	3.27E-08	-0.436026	-2.58E-13
P8_6	-0.412878	2.78E-43	-0.412878	-2.55E-12

Table 8.7: Comparison of Algorithm 2 and Algorithm 2L for solution recovery for eight nonconvex polynomial optimization problems in the literature (ref. [171, 172]). Both algorithms were initialized with solutions to a level-(1,1,0) conditional SAGE relaxation, and Algorithm 2L always recovers an optimal solution. It is especially notable that Algorithm 2L recovers optimal solutions for problems P4_6 and P4_8, since level-(1,1,0) relaxations do not produce tight bounds for these problems.

8.5.3   Minimizing random sparse quartics over the sphere

Here we describe how SAGE relaxations fare for minimizing sparse quartic forms over the unit sphere. This class of test problems is inspired from similar computational experiments by Ahmadi and Majumdar in their work on LP and SOCP-based inner-approximations of the SOS cone [92].

Our method for generating these problems is as follows: initialize $f = 0$ as a polynomial in n variables, and proceed to iterate over all tuples "t" in $[n]^4$. With

probability $n \log n / n^4$, sample a coefficient c_t from the standard normal distribution, and add the term $c_t \boldsymbol{x}^{\alpha_t}$ to f, where $\alpha_t \in [4]^n$ has $\alpha_{tj} = |\{i : t_i = j\}|$. The expected number of terms in f after this procedure is roughly $n \log n$. Once a polynomial is generated, we solve a level-$(0,2,0)$ conditional SAGE relaxation for $(f, g)^{\star}_{\mathbb{R}^n}$, where $g(\boldsymbol{x}) = 1 - \boldsymbol{x}^\top \boldsymbol{x}$.[3] The set "X" in the conditional SAGE relaxation is $\mathsf{X} = \{\boldsymbol{x} : g(\boldsymbol{x}) \geq 0\}$. Figure 8.1 and Table 8.8 report results for 20 problems in 10 variables, 20 problems in 20 variables, 14 problems in 30 variables, and 10 problems in 40 variables.

Figure 8.1: Upper-bounds on the optimality gap $|(f, g)^{(0,2,0)}_{\mathsf{X}} - (f, g)^{\star}_{\mathbb{R}^n}| / |(f, g)^{\star}_{\mathbb{R}^n}|$. The value $(f, g)^{\star}_{\mathbb{R}^n}$ in these calculations was replaced by the objective value of a solution produced by Algorithm 2L. SAGE solved 4 problems in 10 variables, 10 problems in 20 variables, 6 problems in 30 variables, and 4 problems in 40 variables.

8.6 Broader observations from numerical experiments

Here we provide expanded remarks on three aspects of our numerical experiments. First, we address which SAGE relaxations appear to be numerically difficult, and provide a reason for why this might happen. Then we report some of the sizes of the SAGE relaxations as represented by cone programs suitable for low-level solvers. Finally, we demonstrate that these low-level cone programs actually have an extremely useful substructure which is not exploited by MOSEK or ECOS.

---

[3]Because f is homogeneous, $\boldsymbol{x}^\top \boldsymbol{x} = 1$ may be relaxed to $\boldsymbol{x}^\top \boldsymbol{x} \leq 1$ without loss of generality.

solve time (s)	$n = 10$	$n = 20$	$n = 30$	$n = 40$
mean	7.54E−01	6.50E−00	6.46E+01	4.59E+02
std dev.	8.74E−02	8.54E−01	1.38E+01	7.20E+01

Table 8.8: Solver runtimes for level-(0,2,0) conditional SAGE relaxations, on Machine **W**. Similar runtimes can be expected for Machine **L** with $n \in \{10, 20, 30\}$. Solve times with Machine **L** can take much longer for $n \geq 40$, since only part of the problem fits in RAM.

For both signomial and polynomial optimization problems, there is significant benefit to solving level-(p, q, ℓ) relaxations with Lagrange multiplier complexity $p > 0$. However, we encountered several problems where any choice of $p > 0$ resulted in a solver failure due to numerical issues. One explanation for the difficulty of such SAGE relaxations could be how (8.4) and (8.5) set the complexity of a Lagrange multiplier without consideration to its associated constraint function. Compare to the usual SOS-based Lasserre hierarchy, where a degree k constraint polynomial $g_i(x) \geq 0$ appearing in a degree $2d$ relaxation gets an SOS multiplier of degree $2\lfloor (2d - k)/2 \rfloor$. For SAGE relaxations, one could set the support of a Lagrange multiplier with consideration to how the product of the constraint function and Lagrange multiplier affect the sparsity pattern of the final Lagrangian. Suitably chosen supports for nonconstant Lagrange multipliers could also result in bounds which are stronger than those produced by reference hierarchies (8.4) and (8.5).

Remark 8.6.1. The hierarchy for signomial optimization in (6.4.1) addresses the problem mentioned above!

One of the main functions of `sageopt` is to cast abstract SAGE relaxations into low-level standard forms, with feasible sets $\{x \in \mathbb{R}^d : Ax + b \in K\}$ for some $A \in \mathbb{R}^{N \times d}$ and $K \subset \mathbb{R}^N$ which is a product of elementary convex cones. There are several settings within `sageopt` which affect how this compilation process is performed. The impact of different settings for the use of slack variables becomes very apparent as one solves SAGE relaxations farther up a hierarchy (Table 8.9). Regardless of compilation settings, the resulting cone programs end up being large and sparse as reference hierarchy parameters increase. It is possible to construct smaller cone programs by inferring signs on certain coefficients of a modulated Lagrangian, and then appealing to Corollary 4.2.2. `Sageopt` already performs a simple version of such dimension-reduction, which is particularly helpful for the minimax-free hierarchy defined in Equation 4.9.

Level (p, q, ℓ)	(0,1,0)	(1,1,0)	(1,1,1)
(N, d, s): slacks	(953, 342, 4E-3)	(80542, 21896, 8E-5)	(2275860, 574775, 3E-6)
(N, d, s): direct	(840, 229, 6E-3)	(62653, 4007, 5E-4)	(1714437, 13352, 1E-4)

Table 8.9: Dimensions of $A \in \mathbb{R}^{N \times d}$ and sparsity $s = \text{nnz}(A)/(Nd)$ in `sageopt`'s low-level representation of feasible sets for dual relaxations of [171, Problem P4_6]. "Slacks" is the default behavior for `sageopt` version 0.2, which was used for experiments in this article. "Direct" (no slacks) is the default for `sageopt` version 0.5.2. See also Tables 8.6 and 8.7.

structure in SAGE relaxations: sparse cholesky factors of order-d positive definite KKT systems

$(p, q, \ell) = (0, 1, 0)$ $(p, q, \ell) = (1, 1, 0)$ $(p, q, \ell) = (1, 1, 1)$
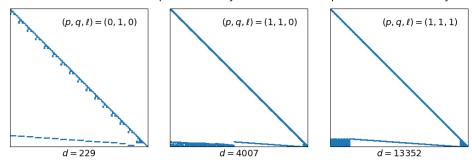
$d = 229$ $d = 4007$ $d = 13352$

Figure 8.2:   Sparsity patterns of Cholesky factors, which can be used to solve linearized KKT systems in interior-point methods for SAGE relaxations to [171, Problem P4_6]. All Cholesky factors used a simple reversed elimination order $d, d-1, \dots, 1$ relative to the original positive definite KKT systems which follow from `sageopt` 0.5.2's low-level problem data.

Solving linearized KKT equations is the largest computational expense in each iteration of an interior-point algorithm for convex cone programming. For ease of exposition, suppose equality constraints in a low-level cone program are represented by two-sided inequality constraints. Under this assumption, both MOSEK and ECOS solve linearized KKT equations by performing a sparse LDL factorization of a symmetric indefinite matrix of order $N + d$. Another approach to solving the linearized KKT equations applies a block-elimination to the indefinite system, to obtain a symmetric positive definite system of dimension $d \ll N$ [177]. In the case of cone programs generated by `sageopt` for dual SAGE relaxations, the only coupling across dual AGE cones occurs through the moment vector v, therefore reduction to positive definite KKT systems would be extremely efficient (see Figure 8.2). It is of interest to see how SAGE relaxations scale with a solver which could take advantage of this structure. The Julia-based `Hypatia.jl` convex optimization solver is a good candidate for such a task [178].

A p p e n d i x   A

PROOFS FOR CHAPTER 3

A.1 Proof of Lemma 3.3.6

Denote $\mathrm{supp}_B(\lambda) = \{b_j : j \in [d], \lambda_j \neq 0\}$. The proof is constructive, where there is nothing to prove when $\mathrm{supp}_B \lambda$ is simplicial. Suppose then that $\lambda \in \Lambda$ has *nonsimplicial* $\mathrm{supp}_B(\lambda)$. We show that it is possible to decompose $\lambda = z\lambda^{(1)} + (1 - z)\lambda^{(2)}$ for some $z \in (0, 1)$ and $\lambda^{(i)} \in \Lambda$ where $\mathrm{supp}\,\lambda^{(i)} \subsetneq \mathrm{supp}\,\lambda$. It should be clear that if this is possible, then the process may be continued in a recursive way if either $\mathrm{supp}_B(\lambda^{(i)})$ are nonsimplicial, and so the claim would follow.

The statement "$\lambda \in \Lambda$" means that h may be expressed as a convex combination of vectors in $\mathrm{supp}_B(\lambda)$, and so by Minkowski-Carathéodory, there exists at least one $\lambda^{(1)}$ in Λ_x with $\mathrm{supp}\,\lambda^{(1)} \subsetneq \mathrm{supp}\,\lambda$ and simplicial $\mathrm{supp}_B(\lambda^{(1)})$. We will use λ and $\lambda^{(1)}$ to construct the desired $\lambda^{(2)}$ and z.

For each real t, consider $\lambda'_t := \lambda^{(1)} + t(\lambda - \lambda^{(1)})$. It is easy to see that for all t, the vector λ'_t belongs to the affine subspace $\{w : h = Bw, \mathbf{1}^\top w = 1\}$, and furthermore the support of λ'_t is contained within the support of λ. Now define $T = \max\{t : \lambda'_t \text{ in } \Delta_d\}$; we claim that $T > 1$ and that the support of λ'_T is a proper subset of the support of λ. The latter claim is more or less immediate. To establish the former claim consider how λ'_t (as an affine combination of $\lambda^{(1)}, \lambda$) belongs to Δ_d if and only if it is elementwise nonnegative. This lets us write $T = \max\{t : \lambda'_t \geq \mathbf{0}\}$. Next, use our knowledge about the support of λ'_t to rewrite the constraint "$\lambda'_t \geq \mathbf{0}$" as "$\lambda_i^{(1)} + t(\lambda_i - \lambda_i^{(1)}) \geq 0$ for all i in $\mathrm{supp}\,\lambda$." Once written in this form, we see that for $t = 1$ all constraints are satisfied strictly. It follows that $T > 1$ at optimality and that the support of λ'_T is distinct from (read: *a proper subset of*) that of λ.

We complete the proof by setting $\lambda^{(2)} = \lambda'_T$ and $z = 1 - 1/T$.

A.2 Proof of Lemma 3.4.4

Denote $f = \mathrm{Sig}(\mathcal{A}, c)$ and $g = \mathrm{Sig}_\mathsf{F}(\mathcal{A}, c)$ for the face F of $\mathsf{P} := \mathrm{conv}(\mathcal{A})$. We may assume without loss of generality that P contains the origin. If $\mathsf{F} = \mathsf{P}$ then $g = f$ and the claim is trivial. If otherwise, the affine hull of F must have some positive codimension ℓ, and there exist supporting hyperplanes $\{\mathsf{S}_i\}_{i=1}^{\ell}$ such that $\mathsf{F} = [\cap_{i=1}^{\ell} \mathsf{S}_i] \cap \mathsf{P}$. We can express S_i as $\{x : \langle s_i, x \rangle = r_i\}$ for a vector s_i and a

scalar r_i. Because P is convex we know that it is contained in one of the half spaces $\{x : \langle s_i, x \rangle \leq r_i\}$ or $\{x : \langle s_i, x \rangle \geq r_i\}$. By possibly replacing (s_i, r_i) by $(-s_i, -r_i)$, we can assume that P is contained in $\{x : \langle s_i, x \rangle \leq r_i\}$. In addition, the assumption that $\mathbf{0}$ belongs to P ensures that each r_i is nonnegative. Now define $s = \sum_{i \in [\ell]} s_i$ and $r = \sum_{i \in [\ell]} r_i \geq 0$. The pair (s, r) is constructed to satisfy the following properties:

- For every α in F, we have $\langle \alpha, s \rangle = r$.

- For every α *not* in F, we have $\langle \alpha, s \rangle < r$.

Finally, define $h = f - g$, which has no exponents in F. The remainder of the proof is case analysis on r.

If $r = 0$ then we must have $r_i = 0$ for all i. The condition that $r_i = 0$ for all i implies that F is contained in a linear subspace U which is orthogonal to s, and so nonnegativity of g over $\mathbb{R}^n$ reduces to nonnegativity of g over U. Suppose then that there exists some $\hat{x}$ in U where $g(\hat{x})$ is negative. For any vector y in the orthogonal complement of U we have $g(\hat{x} + y) = g(\hat{x})$. Meanwhile no matter the value of $\hat{x}$ we know that $\lim_{t \to \infty} h(\hat{x} + ts) = 0$. Using $f_{\mathbb{R}^n}^\star \leq \inf\{f(\hat{x} + ts) : t \text{ in } \mathbb{R}\} \leq g(\hat{x})$, we have the desired result for $r = 0$: $g^\star < 0$ implies $f_{\mathbb{R}^n}^\star < 0$.

Now consider the case when r is positive. Define the vector $\hat{s} = rs / \|s\|^2$; we produce an upper bound on $f_{\mathbb{R}^n}^\star$ by searching over all hyperplanes $\{x : \langle \hat{s}, x \rangle = t\}$ for t in $\mathbb{R}$. Specifically, for any x in $\mathbb{R}^n$ there exists a scalar t and a vector y such that $x = t\hat{s} + y$ and $\langle \hat{s}, y \rangle = 0$. In these terms we have

$$g(t\hat{s} + y) = \exp(t\|\hat{s}\|^2) \sum_{\alpha \in \mathcal{A} \cap F} c_\alpha \underbrace{\exp(t\langle \alpha - \hat{s}, \hat{s} \rangle)}_{=1 \text{ for all } t} \exp\langle \alpha, y \rangle. \tag{A.1}$$

Hence assuming $g^\star < 0$ means $\sum_{\alpha \in \mathcal{A} \cap F} c_\alpha \exp\langle \alpha, \hat{y} \rangle < 0$ for some $\hat{y}$ in $\mathrm{Span}(s)^\perp$. Using this $\hat{y}$, one may verify that

$$\lim_{t \to \infty} f(t\hat{s} + \hat{y}) = -\infty, \tag{A.2}$$

and so when r is positive, $g^\star < 0$ implies $f_{\mathbb{R}^n}^\star < 0$.

A.3   Proof of Proposition 3.6.8

The cases $(2) \Rightarrow (1)$ and $(3) \Rightarrow (1)$ are easy.

$\neg(2) \Rightarrow \neg(1)$. Because $C_{\mathsf{NNS}}(\mathcal{A})$ and $C_{\mathsf{SAGE}}(\mathcal{A})$ are full dimensional closed convex sets, the condition $C_{\mathsf{SAGE}}(\mathcal{A}) \neq C_{\mathsf{NNS}}(\mathcal{A})$ implies that $C_{\mathsf{NNS}}(\mathcal{A}) \setminus C_{\mathsf{SAGE}}(\mathcal{A})$ has

nonempty interior. Assuming this condition, fix a vector $\tilde{c}$ and a radius r such that $B(\tilde{c}, r) \subset C_{\mathsf{NNS}}(\mathcal{A}) \setminus C_{\mathsf{SAGE}}(\mathcal{A})$.[1] This allows us to strictly separate $\tilde{c}$ from $C_{\mathsf{SAGE}}(\mathcal{A})$, which establishes $f_{\mathbb{R}^n}^{\star} \geq f_{\mathbb{R}^n}^{\mathsf{SAGE}} + r > f_{\mathbb{R}^n}^{\mathsf{SAGE}}$.

$(1) \Rightarrow (3)$. Now suppose that $f_{\mathbb{R}^n}^{\star} = f_{\mathbb{R}^n}^{\mathsf{SAGE}}$ for all relevant f. In this case, the function $c \mapsto \inf\{\langle c, v \rangle : v \in \Omega\}$ is the same for $\Omega = \mathrm{cl}\,\mathrm{conv}\,\mathrm{exp}\,\mathcal{A}\mathbb{R}^n$ or $\Omega = \{v : v_1 = 1 \text{ and } v \text{ in } C_{\mathsf{SAGE}}(\mathcal{A})^{\star}\}$. This function completely determines the set of all half spaces containing Ω. Since Ω is closed and convex, it is precisely equal to the intersection of all half spaces containing it; the result follows.

---

[1] $B(x, d)$ is ℓ_2 ball centered at x of radius d.

A p p e n d i x B

PROOFS FOR CHAPTER 5

B.1   A proposition for use in Lemma 5.5.10

Proposition B.1.1. *Suppose* $S \subset \mathbb{R}^m \setminus \{0\}$ *is compact (not necessarily convex) and set* $T = \mathrm{co}\, S$. *If it is known* a priori *that* $\mathrm{cl}\, T$ *is pointed, then* $T = \mathrm{cl}\, T$ *is closed.*

Proof. Since $\mathrm{cl}\, T$ is pointed, there exists a distinguished element $t^\star \in T$ for which $\langle t^\star, t \rangle > 0$ for all $t \in (\mathrm{cl}\, T) \setminus \{0\}$. Consider the set $H = \{t \in T : \langle t^\star, t \rangle = 1\}$. It is clear that H is bounded, $\mathrm{co}\, H = T$, and $0 \notin H$. If H is closed, then by [65, Corollary 9.6.1] we will have that $\mathrm{co}\, H = T$ is also closed. We show that H is closed by directly considering sequences in H. We express these sequences with the help of the m-fold Cartesian product $S^m = S \times \cdots \times S \subset \mathbb{R}^{m^2}$.

Let $(\boldsymbol{h}^{(k)})_{k \in \mathbb{N}} \subset H$ have a limit in $\mathbb{R}^m$. Since H is of dimension at most $m - 1$ and is generated by S, Carathéodory's Theorem tells us that for every k there exists a weighting vector $\boldsymbol{\lambda}^{(k)} \in \mathbb{R}^m_+$ and a block vector $\boldsymbol{q}^{(k)} = (\boldsymbol{s}_1^{(k)}, \ldots, \boldsymbol{s}_m^{(k)}) \in S^m$ where

$$\boldsymbol{h}^{(k)} = \textstyle\sum_{i=1}^m \lambda_i^{(k)} \boldsymbol{s}_i^{(k)}.$$

Because S is compact, the continuous function $s \mapsto \langle t^\star, s \rangle$ attains a minimum on $s^\star \in S$. Since S does not contain zero, we have that $\langle t^\star, s^\star \rangle = a > 0$. It follows that $\lambda_i^{(k)} \leq 1/a$ for each $i \in [m]$ and $K \in \mathbb{N}$. The sequences $(\boldsymbol{\lambda}^{(k)})_{k \in \mathbb{N}} \subset [0, 1/a]^m$ and $(\boldsymbol{q}^{(k)})_{k \in \mathbb{N}} \subset S^m$ are bounded, and therefore $((\boldsymbol{\lambda}^{(k)}, \boldsymbol{q}^{(k)}))_{k \in \mathbb{N}}$ has a convergent subsequence. The limits $\boldsymbol{\lambda}^{(\infty)}$ and $\boldsymbol{q}^{(\infty)}$ of these convergent subsequences must belong to $[0, 1/a]^m$ and S^m, respectively. By continuity, we have

$$\boldsymbol{h}^{(\infty)} := \lim_{k \to \infty} \boldsymbol{h}^{(k)} = \textstyle\sum_{i=1}^m \lambda_i^{(\infty)} \boldsymbol{s}_i^{(\infty)},$$

hence $\boldsymbol{h}^{(\infty)} \in H$. Since we have shown that all convergent sequences in H converge to a *point in* H, we have that H is closed. $\square$

B.2   Proof of Corollary 5.4.5

By Theorem 5.3.7, polyhedral X have finitely many X-circuits, up to scaling. Apply Theorem 5.4.2 and finiteness of the normalized circuits $\Lambda_X(\mathcal{A})$ to write

$$C_X(\mathcal{A}) = \sum_{\lambda \in \Lambda_X(\mathcal{A})} C_X(\mathcal{A}, \lambda).$$

The first claim follows as each of the finitely many sets $C_X(\mathcal{A}, \lambda)$ appearing in the above sum are (dual) power cone representable. For the second claim observe that under the rationality assumptions we have $\Lambda_X(\mathcal{A}) \subset \mathbb{Q}^{\mathcal{A}}$. Using $\beta := \lambda^-$ and $m := |\operatorname{supp}\lambda|$, it is known that the m-dimensional λ-weighted power cone (and its dual) are second-order representable when $\lambda_{\backslash\beta}$ is a rational vector in the $(m-1)$-dimensional probability simplex [142, Section 3.4]. The last claim follows as the semidefinite extension degree of the second-order cone is two [142, Section 2.3].

B.3 Proof of Proposition 5.4.6

First, we note that some inequality constraints $c_{\backslash\beta} \geq 0$ are implied by $(S_\lambda c - r\delta_\beta) \in \operatorname{Pow}(\lambda)^\dagger$. It is necessary to include the inequality constraints explicitly, to account for the case when $\operatorname{supp}\lambda \subsetneq \mathcal{A}$. The condition $(S_\lambda c - r\delta_\beta) \in \operatorname{Pow}(\lambda)^\dagger$ can be rewritten as

$$\prod_{\alpha \in \lambda^+} [c_\alpha/\lambda_\alpha]^{\lambda_\alpha} \geq |c_\beta \exp(\sigma_X(-\mathcal{A}^\dagger\lambda)) - r|. \tag{B.1}$$

Meanwhile, the minimum of $|c_\beta \exp(\sigma_X(-\mathcal{A}^\dagger\lambda)) - r|$ over $r \geq 0$ is attained at $r = 0$ when $c_\beta < 0$ and $r = c_\beta \exp(\sigma_X(-\mathcal{A}^\dagger\lambda))$ when $c_\beta \geq 0$. In the $c_\beta < 0$ case the constraint (B.1) becomes

$$\prod_{\alpha \in \lambda^+} [c_\alpha/\lambda_\alpha]^{\lambda_\alpha} \geq -c_\beta \exp(\sigma_X(-\mathcal{A}^\dagger\lambda)).$$

In the $c_\beta \geq 0$ case the constraint (B.1) is vacuous, since $\prod_{\alpha \in \lambda^+} [c_\alpha/\lambda_\alpha]^{\lambda_\alpha} \geq 0$ is implied by $c_{\backslash\beta} \geq 0$. As the constraint in the preceding display is similarly vacuous when $c_\beta > 0$, we see that it can be used in lieu of (B.1) without loss of generality.

B.4 Proof of Proposition 5.4.7

Let $\beta = \lambda^-$ as is usual. To $v \in \mathbb{R}^{\mathcal{A}}$ associate $\operatorname{Val}(v) = \inf\{v^\top c : c \in C_X(\mathcal{A}, \lambda)\}$. A vector v belongs to $C_X(\mathcal{A}, \lambda)^\dagger$ if and only if $\operatorname{Val}(v) = 0$. We will find constraints on v so that the dual feasible set for computing $\operatorname{Val}(v)$ is nonempty, which in turn will imply $\operatorname{Val}(v) = 0$.

We begin by noting that for any element $\alpha \in \mathcal{A} \setminus \operatorname{supp}\lambda$, the only constraints on c_α, v_α for $c \in C_X(\mathcal{A}, \lambda), v \in C_X(\mathcal{A}, \lambda)^\dagger$ are $c_\alpha \geq 0, v_\alpha \geq 0$; therefore we assume $\mathcal{A} = \operatorname{supp}\lambda$ for the remainder of the proof. When considering the given expression for $\operatorname{Val}(v)$ as a primal problem, we compute a dual using (5.7) from Proposition 5.4.6. Under the assumption $\mathcal{A} = \operatorname{supp}\lambda$, the constraint $c_{\backslash\beta} \geq 0$ is implied by $(S_\lambda c - r\delta_\beta) \in \operatorname{Pow}(\lambda)^\dagger$. Therefore when forming a Lagrangian for $\operatorname{Val}(v)$ using (5.7), the dual variable to "$c_{\backslash\beta} \geq 0$" may be omitted.

For the remaining constraints $(S_\lambda c - r\delta_\beta) \in \text{Pow}(\lambda)^\dagger$ and $r \geq 0$ we use dual variables $\mu \in \text{Pow}(\lambda)$ and $t \in \mathbb{R}_+$ respectively; the Lagrangian is

$$\mathcal{L}(c, r, \mu, t) = v^\top c - \mu^\top (S_\lambda c - r\delta_\beta) - tr$$
$$= c^\top (v - S_\lambda \mu) - r(t - \mu_\beta).$$

For the Lagrangian to be bounded below over $c \in \mathbb{R}^{\mathcal{A}}$ and $r \in \mathbb{R}$, it is necessary and sufficient that $v = S_\lambda \mu$ and $\mu_\beta = t$. Since we have assumed $\text{supp } \lambda = \mathcal{A}$ and $\sigma_X(-\mathcal{A}^\dagger \lambda) < \infty$, the diagonal linear operator S_λ is symmetric positive definite, so we can express the requirements on μ, t as

$$S_\lambda^{-1} v = \mu \quad \text{and} \quad \mu_\beta = t.$$

Therefore the conditions $S_\lambda^{-1} v \in \text{Pow}(\lambda)$, $v_\beta \geq 0$ are equivalent to

$$\text{Val}(v) = \inf \left\{ \sup\{\mathcal{L}(c, r, \mu, t) \ : \ (\mu, t) \in \text{Pow}(\lambda) \times \mathbb{R}_+\} \ : \ (c, r) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R} \right\}$$
$$= \sup \left\{ \inf\{\mathcal{L}(c, r, \mu, t) \ : \ (c, r) \in \mathbb{R}^{\mathcal{A}} \times \mathbb{R}\} \ : \ (\mu, t) \in \text{Pow}(\lambda) \times \mathbb{R}_+ \right\} = 0.$$

The proposition follows by applying the definitions of $\text{Pow}(\lambda)$ and S_λ.

ALGORITHMS FOR CHAPTER 7

In the discussion that follows, $\mathcal{A} = (\alpha_{ij} \; : \; (i, j) \in [m] \times [n])$ is an $m \times n$ matrix with rows α_i.

Algorithm 3 magnitude recovery for dual SAGE polynomial relaxations.

Input: A matrix $\mathcal{A} \in \mathbb{N}^{m \times n}$. Vectors $v \in C_{\mathsf{SAGE}}^{\mathsf{POLY}}(\mathcal{A}, \mathsf{X})^{\dagger}$ and $\hat{v} \in C_{\mathsf{SAGE}}(\mathcal{A}, \mathsf{Y})$. Zero threshold parameter $\epsilon_0 > 0$.

 1: **procedure** VARIABLEMAGNITUDES($\mathcal{A}, v, \hat{v}, \epsilon_0$)
 2: $\quad$ M $\leftarrow$ []
 3: $\quad$ **for** $j = 1, \ldots, m$ **do**
 4: $\quad\quad$ **if** $\hat{v}_j = 0$ **then**
 5: $\quad\quad\quad$ **Continue**
 6: $\quad\quad$ Recover z_j in $\mathbb{R}^n$ s.t. $\hat{v}_j \log(\hat{v}/\hat{v}_j) \geq [\mathcal{A} - \mathbf{1}\alpha_j]z_j$ and $(z_j, \hat{v}_j) \in \mathrm{co}\,\mathsf{Y}$.
 7: $\quad\quad$ M.append($\exp(z_j/\hat{v}_j)$)
 8: $\quad$ **if** $(x^{\alpha_1}, \ldots, x^{\alpha_m}) \neq |v|$ for all x in M **then**
 9: $\quad\quad$ Compute (y, t) solving Problem 7.5, for given ϵ_0.
10: $\quad\quad$ M.append($\exp y$)
11: $\quad$ **return** M.

As in the signomial case, Algorithm 3 always returns a vector $x \in \mathsf{X}$. Assuming that z from Line 7 are already computed as part of representing $\hat{v}$, the complexity of this algorithm is dominated by Line 12. The runtime of Line 12 is in turn negligible relative to solving a SAGE relaxation to obtain vectors v and $\hat{v}$. Infeasibility errors encountered in Line 12 should be handled by jumping to Line 15.

Algorithm 4 sign recovery for dual SAGE polynomial relaxations.

Input: A matrix $\mathcal{A} \in \mathbb{N}^{m \times n}$. A vector v in $\mathbb{R}^m$. A Boolean heuristic.

1: **procedure** VARIABLESIGNS($\mathcal{A}, v$, heuristic)
2:      $\mathsf{U} \leftarrow \{i : v_i \neq 0 \text{ and } \alpha_i \text{ is not even }\}$
3:      $\mathsf{W} \leftarrow \{j : \alpha_{ij} \equiv 1 \mod 2 \text{ for some } i \text{ in } \mathsf{U}\}$
4:      $\mathsf{Z} \leftarrow \{z \in \{0,1\}^n : \mathcal{A}[\mathsf{U},:]z \equiv (v < 0)[\mathsf{U}] \mod 2, z_i = 0 \text{ for } i \text{ in } [n]\backslash\mathsf{W}\}$
5:      $\mathsf{S} \leftarrow \{\}$
6:      **for** z in Z **do**
7:          $s \leftarrow \mathbf{1}$
8:          **for** j in $\{j : \alpha_{ij} > 0 \text{ for some } i \text{ in } \mathsf{U}\}$ **do**
9:              $s_j \leftarrow -1 \text{ if } z_j = 1, \ 1 \text{ if } z_j = 0$
10:          $\mathsf{S} \leftarrow \mathsf{S} \cup \{s\}$
11:      If $\mathsf{S} = \emptyset$ and heuristic, update $\mathsf{S} \leftarrow \{\text{HueristicSigns}(\mathcal{A}, v)\}$.
12:      **return** S.

Let us describe the ways in which Algorithm 4 differs from the discussion in Subsection 7.3.2. First- there are changes to the sets U and W. The set U now drops any rows α_i from $\mathcal{A}$ where α_i is even; it is easy to verify that this does not affect the set of solutions to the appropriate linear system. The set W changes by only considering j where at least one $\alpha_{ij} \equiv 1 \mod 2$. This change is valid because if α_{ij} is even for all i, then the sign of variable x_j is irrelevant to the underlying optimization problem, and we make take $x_j \geq 0$ without loss of generality.

Next we speak to the "hueristic" sign recovery. We partly mean to leave this as open-ended, however for completeness we describe the algorithm used in `sageopt`. The goal is to find a vector s in $\{+1, -1\}$ so that the signs of $s^{\mathcal{A}} \doteq (s^{\alpha_1}, \ldots, s^{\alpha_m})$ match the signs of v to the greatest extent possible. However, we consider how having s^{α_i} match the sign of v_i may not be very important if v_i is very small. Therefore we use a merit function $M(s) = v^{\top}s^{\mathcal{A}}$ to evaluate the quality of candidate signs s. We apply a greedy algorithm to maximize the merit function $M(s)$ as follows: initialize $s = \mathbf{1}$, and a set of undecided coordinates $\mathsf{C} = \{1, \ldots, n\}$. As long as the set C is nonempty, find an index $i^\star \in \mathsf{C}$ so that changing $s_{i^\star} = 1$ to $s_{i^\star} = -1$ maximizes improvement in the merit function. If the improvement is positive, then perform the update $s_{i^\star} \leftarrow -1$. Regardless of whether or not the improvement is positive, remove $i^\star$ from C. Once C is empty, return s.

BIBLIOGRAPHY

[1] N. Z. Shor. "Class of global minimum bounds of polynomial functions". In: *Cybernetics* 23.6 (Nov. 1987), pp. 731–734. DOI: `10.1007/bf01070233`.

[2] Pablo Parrilo. "Structured semidefinite programs and semialgebraic geometry methods in robustness and optimization". PhD thesis. California Institute of Technology, May 2000.

[3] Jean B. Lasserre. "Global Optimization with Polynomials and the Problem of Moments". In: *SIAM Journal on Optimization* 11.3 (Jan. 2001), pp. 796–817. DOI: `10.1137/s1052623400366802`.

[4] A. Papachristodoulou and S. Prajna. "On the construction of Lyapunov functions using the sum of squares decomposition". In: *Proceedings of the 41st IEEE Conference on Decision and Control, 2002*. Vol. 3. 2002, 3482–3487 vol.3.

[5] David Papp. "OPTIMIZATION MODELS FOR SHAPE-CONSTRAINED FUNCTION ESTIMATION PROBLEMS INVOLVING NONNEGATIVE POLYNOMIALS AND THEIR RESTRICTIONS". PhD thesis. Rutgers, The State University of New Jersey, May 2011.

[6] Mihaela Curmei and Georgina Hall. *Shape-Constrained Regression using Sum of Squares Polynomials*. 2020. arXiv: `2004.03853 [math.OC]`.

[7] Martin Mevissen et al. "Solving partial differential equations via sparse SDP relaxations". In: *Pacific Journal of Optimization* 4.2 (2008), pp. 213–241.

[8] Milan Korda, Didier Henrion, and Jean-Bernard Lasserre. *Moments and convex optimization for analysis and control of nonlinear partial differential equations*. 2018. arXiv: `1804.07565 [math.OC]`.

[9] S. Marx et al. "A moment approach for entropy solutions to nonlinear hyperbolic PDEs". In: *Mathematical Control & Related Fields* 10.1 (2020), pp. 113–140. DOI: `10.3934/mcrf.2019032`. URL: `https://doi.org/10.3934/mcrf.2019032`.

[10] Konrad Schmüdgen. "The K-moment problem for compact semi-algebraic sets". In: *Mathematische Annalen* 289.1 (1991), pp. 203–206.

[11] Mihai Putinar. "Positive Polynomials on Compact Semi-algebraic Sets". In: *Indiana University Mathematics Journal* 42.3 (1993), pp. 969–984. ISSN: 00222518, 19435258. URL: `http://www.jstor.org/stable/24897130`.

[12] Grigoriy Blekherman. "There are significantly more nonnegative polynomials than sums of squares". In: *Israel Journal of Mathematics* 153.1 (Dec. 2006), pp. 355–380. DOI: `10.1007/bf02771790`. URL: `https://doi.org/10.1007/bf02771790`.

[13] V. Chandrasekaran and P. Shah. "Relative entropy relaxations for signomial optimization". In: *SIAM J. Optim.* 26.2 (2016), pp. 1147–1173. DOI: `10.1137/140988978`.

[14] Bruce Reznick. "Forms Derived from the Arithmetic-Geometric Inequality." In: *Mathematische Annalen* 283.3 (1989), pp. 431–464.

[15] S. Müller et al. "Sign Conditions for Injectivity of Generalized Polynomial Maps with Applications to Chemical Reaction Networks and Real Algebraic Geometry". In: *Found. Comp. Math.* 16.1 (2015), pp. 69–97. DOI: `10.1007/s10208-014-9239-3`. URL: `https://doi.org/10.1007/s10208-014-9239-3`.

[16] S. Müller, J. Hofbauer, and G. Regensburger. "On the Bijectivity of Families of Exponential/Generalized Polynomial Maps". In: *SIAM J. Appl. Algebra and Geometry* 3.3 (2019), pp. 412–438. DOI: `10.1137/18m1178153`. URL: `https://doi.org/10.1137/18m1178153`.

[17] Casian Pantea, Heinz Koeppl, and Gheorghe Craciun. "Global injectivity and multiple equilibria in uni- and bi-molecular reaction networks". In: *Discrete and Continuous Dynamical Systems - Series B* 17.6 (May 2012), pp. 2153–2170. DOI: `10.3934/dcdsb.2012.17.2153`. URL: `https://doi.org/10.3934/dcdsb.2012.17.2153`.

[18] J. Silipo. "The Ronkin number of an exponential sum". In: *Mathematische Nachrichten* 285.8-9 (2012), pp. 1117–1129. DOI: `10.1002/mana.201000130`. URL: `https://doi.org/10.1002/mana.201000130`.

[19] J. Forsgård and T. de Wolff. *The Lattice of Amoebas*. 2017. eprint: `arXiv:1711.02705`.

[20] A. A Ergür, G. Paouris, and J. M. Rojas. "Tropical varieties for exponential sums". In: *Math. Annalen* 377 (2020), pp. 863–882.

[21] J. Forsgård and T. de Wolff. *The algebraic boundary of the SONC cone*. Preprint, arXiv:1905.04776. 2019.

[22] M. Dressler et al. "Global Optimization via the Dual SONC Cone and Linear Programming". In: *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC)*. ACM, 2020, pp. 138–145.

[23] Philippe Moustrou et al. *Symmetry reduction in AM/GM-based optimization*. 2021. arXiv: `2102.12913 [math.OC]`.

[24] Stephen Boyd et al. "A tutorial on geometric programming". In: *Optimization and engineering* 8.1 (2007), p. 67.

[25] F. Horn and R. Jackson. "General mass action kinetics". In: *Archive for Rational Mechanics and Analysis* 47.2 (1972), pp. 81–116. DOI: `10.1007/bf00251225`. URL: `https://doi.org/10.1007/bf00251225`.

[26] E. Voit, H. Martens, and S. Omholt. "150 Years of the Mass Action Law". In: *PLoS Computational Biology* 11.1 (2015). Ed. by Ruth Nussinov, e1004012. DOI: `10.1371/journal.pcbi.1004012`. URL: `https://doi.org/10.1371/journal.pcbi.1004012`.

[27] G. Craciun and M. Feinberg. "Multiple Equilibria in Complex Chemical Reaction Networks: I. The Injectivity Property". In: *SIAM Journal on Applied Mathematics* 65.5 (Jan. 2005), pp. 1526–1546. DOI: `10.1137/s0036139904440278`. URL: `https://doi.org/10.1137/s0036139904440278`.

[28] G. Craciun and M. Feinberg. "Multiple Equilibria in Complex Chemical Reaction Networks: II. The Species-Reaction Graph". In: *SIAM Journal on Applied Mathematics* 66.4 (Jan. 2006), pp. 1321–1338. DOI: `10.1137/050634177`. URL: `https://doi.org/10.1137/050634177`.

[29] Elias August, Gheorghe Craciun, and Heinz Koeppl. "Finding invariant sets for biological systems using monomial domination". In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, Dec. 2012. DOI: `10.1109/cdc.2012.6426491`. URL: `https://doi.org/10.1109/cdc.2012.6426491`.

[30] Elisenda Feliu et al. "The Kinetic Space of Multistationarity in Dual Phosphorylation". In: *Journal of Dynamics and Differential Equations* (Sept. 2020). ISSN: 1572-9222. DOI: `10.1007/s10884-020-09889-6`. URL: `http://dx.doi.org/10.1007/s10884-020-09889-6`.

[31] Polly Y. Yu and Gheorghe Craciun. "Mathematical Analysis of Chemical Reaction Systems". In: *Israel Journal of Chemistry* 58.6-7 (2018), pp. 733–741. DOI: `10.1002/ijch.201800003`. URL: `https://doi.org/10.1002/ijch.201800003`.

[32] Clarence Zener. "A MATHEMATICAL AID IN OPTIMIZING ENGINEERING DESIGNS". In: *Proceedings of the National Academy of Sciences* 47.4 (1961), pp. 537–539. ISSN: 0027-8424. DOI: `10.1073/pnas.47.4.537`. eprint: `https://www.pnas.org/content/47/4/537.full.pdf`. URL: `https://www.pnas.org/content/47/4/537`.

[33] C. Zener. "A FURTHER MATHEMATICAL AID IN OPTIMIZING ENGINEERING DESIGNS". In: *Proceedings of the National Academy of Sciences* 48.4 (Apr. 1962), pp. 518–522. DOI: `10.1073/pnas.48.4.518`. URL: `https://doi.org/10.1073/pnas.48.4.518`.

[34] Richard J. Duffin, Elmor L. Peterson, and Clarence Zener. *Geometric programming. Theory and application*. English. New York-London-Sydney: John Wiley and Sons, Inc. xiii, 278 p. (1967). 1967.

[35] Richard James Duffin and Elmor L Peterson. "Geometric programming with signomials". In: *Journal of Optimization Theory and Applications* 11.1 (1973), pp. 3–35.

[36] K. G. Murty and S. N. Kabadi. "Some NP-complete problems in quadratic and nonlinear programming". In: *Math. Program.* 39.2 (1987), pp. 117–129. DOI: `10.1007/bf02592948`. URL: `https://doi.org/10.1007/bf02592948`.

[37] Yurii Nesterov and Arkadii Nemirovskii. *Interior-Point Polynomial Algorithms in Convex Programming*. Society for Industrial and Applied Mathematics, Jan. 1994. DOI: `10.1137/1.9781611970791`. URL: `https://doi.org/10.1137/1.9781611970791`.

[38] Mung Chiang. *Geometric programming for communication systems*. Now Publishers Inc, 2005.

[39] Gary E. Blau and Douglass J. Wilde. "A Lagrangian algorithm for equality constrained generalized polynomial optimization". In: *AIChE Journal* 17.1 (1971), pp. 235–240. DOI: `https://doi.org/10.1002/aic.690170145`. eprint: `https://aiche.onlinelibrary.wiley.com/doi/pdf/10.1002/aic.690170145`. URL: `https://aiche.onlinelibrary.wiley.com/doi/abs/10.1002/aic.690170145`.

[40] R. Dembo and M. Avriel. "Optimal design of a membrane separation process using signomial programming". In: *Mathematical Programming* 15.1 (Dec. 1978), pp. 12–25.

[41] C. A. Floudas et al. "Generalized Geometric Programming Problems". In: *Handbook of Test Problems in Local and Global Optimization*. Boston, MA: Springer US, 1999, pp. 85–105. ISBN: 978-1-4757-3040-1. DOI: `10.1007/978-1-4757-3040-1_7`. URL: `https://doi.org/10.1007/978-1-4757-3040-1_7`.

[42] K.-M. Björk and T. Westerlund. "Global optimization of heat exchanger network synthesis problems with and without the isothermal mixing assumption". In: *Computers & Chemical Engineering* 26.11 (2002), pp. 1581–1593.

[43] M. Avriel and J. D. Barrett. "Optimal Design of Pitched Laminated Wood Beams". In: *Advances in Geometric Programming*. Springer US, 1980, pp. 407–419. DOI: `10.1007/978-1-4615-8285-4_21`. URL: `https://doi.org/10.1007/978-1-4615-8285-4_21`.

[44] Hojjat Adeli and Osama Kamal. "Efficient optimization of space trusses". In: *Computers & Structures* 24.3 (1986), pp. 501–511. ISSN: 0045-7949. DOI: `https://doi.org/10.1016/0045-7949(86)90327-5`.

[45] Balaur S. Dhillon and Chen-Hsing Kuo. "Optimum Design of Composite Hybrid Plate Girders". In: *Journal of Structural Engineering* 117.7 (1991), pp. 2088–2098. DOI: `10.1061/(ASCE)0733-9445(1991)117:7(2088)`.

[46] Yun Kang Sui and Xi Cheng Wang. "Second-order method of generalized geometric programming for spatial frame optimization". In: *Computer Methods in Applied Mechanics and Engineering* 141.1-2 (Feb. 1997), pp. 117–123. DOI: `10.1016/s0045-7825(96)01097-3`.

[47] R.A. Jabr. "Inductor design using signomial programming". In: *COMPEL - The international journal for computation and mathematics in electrical and electronic engineering* 26.2 (Apr. 2007). Ed. by R. Martone, pp. 461–475.

[48] H. Kwan et al. "Designing Globally Optimal Delta-Sigma Modulator Topologies via Signomial Programming". In: *2007 International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*. Apr. 2007, pp. 1–4.

[49] Mung Chiang. "Nonconvex Optimization for Communication Networks". In: *Advances in Applied Mathematics and Global Optimization: In Honor of Gilbert Strang*. Boston, MA: Springer US, 2009, pp. 137–196. ISBN: 978-0-387-75714-8. DOI: `10.1007/978-0-387-75714-8_5`.

[50] Berk Ozturk and Ali Saab. "Optimal Aircraft Design Decisions under Uncertainty via Robust Signomial Programming". In: *AIAA Aviation 2019 Forum*. American Institute of Aeronautics and Astronautics, June 2019. DOI: `10.2514/6.2019-3351`. URL: `https://doi.org/10.2514/6.2019-3351`.

[51] Martin A. York, Warren W. Hoburg, and Mark Drela. "Turbofan Engine Sizing and Tradeoff Analysis via Signomial Programming". In: *Journal of Aircraft* 55.3 (May 2018), pp. 988–1003. DOI: `10.2514/1.c034463`. URL: `https://doi.org/10.2514/1.c034463`.

[52] Martin A. York et al. "Efficient Aircraft Multidisciplinary Design Optimization and Sensitivity Analysis via Signomial Programming". In: *AIAA Journal* 56.11 (Nov. 2018), pp. 4546–4561. DOI: `10.2514/1.j057020`. URL: `https://doi.org/10.2514/1.j057020`.

[53] Philippe G. Kirschen and Warren W. Hoburg. "The Power of Log Transformation: A Comparison of Geometric and Signomial Programming with General Nonlinear Programming Techniques for Aircraft Design Optimization". In: *2018 AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. American Institute of Aeronautics and Astronautics, Jan. 2018. DOI: `10.2514/6.2018-0655`. URL: `https://doi.org/10.2514/6.2018-0655`.

[54] Philippe G. Kirschen et al. "Application of Signomial Programming to Aircraft Design". In: *Journal of Aircraft* 55.3 (May 2018), pp. 965–987. DOI: `10.2514/1.c034378`. URL: `https://doi.org/10.2514/1.c034378`.

[55] D. Hall et al. "Assessment of a Boundary Layer Ingesting Turboelectric Aircraft Configuration using Signomial Programming". In: *2018 Aviation*

Technology, Integration, and Operations Conference. American Institute of Aeronautics and Astronautics, June 2018. DOI: `10.2514/6.2018-3973`.

[56] Philippe Kirschen and Edward Burnell. *Hyperloop System Optimization*. 2021. arXiv: `2104.03907 [cs.CE]`.

[57] Snorri Gudmundsson. *General aviation aircraft design: applied methods and procedures*. First edition. OCLC: ocn863648682. Oxford ; Waltham, MA: Butterworth-Heinemann, 2014. ISBN: 9780123973085.

[58] Dr. Don W. Green and Dr. Marylee Z. Southard. *Perry's Chemical Engineers' Handbook, 9th Edition*. en. 9th edition. New York: McGraw-Hill Education, 2019. ISBN: 9780071834087. URL: `https://www.accessengineeringlibrary.com/content/book/9780071834087`.

[59] Riley Murray, Venkat Chandrasekaran, and Adam Wierman. "Newton Polytopes and Relative Entropy Optimization". In: *Foundations of Computational Mathematics* (Mar. 2021). DOI: `10.1007/s10208-021-09497-w`. URL: `https://doi.org/10.1007/s10208-021-09497-w`.

[60] R. Murray, V. Chandrasekaran, and A. Wierman. "Signomial and polynomial optimization via relative entropy and partial dualization". In: *Mathematical Programming Computation* (Oct. 2020). DOI: `10.1007/s12532-020-00193-4`. URL: `https://doi.org/10.1007/s12532-020-00193-4`.

[61] Riley Murray, Helen Naumann, and Thorsten Theobald. *Sublinear circuits and the constrained signomial nonnegativity problem*. 2020. eprint: `arXiv:2006.06811`.

[62] Mareike Dressler and Riley Murray. *An algebraic perspective to signomial optimization*. In preparation. 2021.

[63] Lukas Wachter et al. *A convex relaxation approach for the optimized pulse pattern problem*. 2020. eprint: `arXiv:2010.14853`.

[64] Lukas Wachter. "On Robustness and Smoothness of Optimized Pulse Patterns for Power Conversion Systems". MA thesis. Aalto University, School of Electrical Engineering, 2020.

[65] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1997.

[66] "1. Preliminaries". In: *A Mathematical View of Interior-Point Methods in Convex Optimization*. Society for Industrial and Applied Mathematics, Jan. 2001, pp. 1–20. DOI: `10.1137/1.9780898718812.ch1`. URL: `https://doi.org/10.1137/1.9780898718812.ch1`.

[67] S. Mehrotra. "On the Implementation of a Primal-Dual Interior Point Method". In: *SIAM Journal on Optimization* 2.4 (Nov. 1992), pp. 575–601. DOI: `10.1137/0802028`. URL: `https://doi.org/10.1137/0802028`.

[68] Leonid Faybusovich. In: *Positivity* 1.4 (1997), pp. 331–357. DOI: `10.1023/a:1009701824047`. URL: `https://doi.org/10.1023/a:1009701824047`.

[69] J. Dahl and E. Andersen. *A primal-dual interior-point algorithm for nonsymmetric exponential-cone optimization.* Preprint, `http://www.optimization-online.org/DB_HTML/2019/05/7227.html`. 2019.

[70] Riley Badenbroek and Joachim Dahl. *An Algorithm for Nonsymmetric Conic Optimization Inspired by MOSEK*. 2020. arXiv: `2003.01546` [`math.OC`].

[71] A.G. Khovanskii. "Fewnomials and Pfaff manifolds". In: *Proceedings of the International Congress of Mathematicians*. Vol. 1. 1983.

[72] A.G. Khovanskii. *Fewnomials*. Vol. 88. American Mathematical Soc., 1991.

[73] Bruce Reznick. "Extremal PSD forms with few terms". In: *Duke Mathematical Journal* 45.2 (June 1978), pp. 363–374. DOI: `10.1215/s0012-7094-78-04519-2`.

[74] T. Bajbar and O. Stein. "Coercive Polynomials and Their Newton Polytopes". In: *SIAM Journal on Optimization* 25.3 (Jan. 2015), pp. 1542–1570.

[75] C. Maranas and C. Floudas. "Global optimization in generalized geometric programming". In: *Computers & Chemical Engineering* 21.4 (Dec. 1997), pp. 351–369.

[76] H. Sherali. "Global optimization of nonconvex polynomial programming problems having rational exponents". In: *Journal of Global Optimization* 12.3 (1998), pp. 267–283.

[77] Peiping Shen and Kecun Zhang. "Global optimization of signomial geometric programming using linear relaxation". In: *Applied Mathematics and Computation* 150.1 (Feb. 2004), pp. 99–114. DOI: `10.1016/s0096-3003(03)00200-5`.

[78] Y. Wang and Z. Liang. "A Deterministic Global Optimization Algorithm for Generalized Geometric Programming". In: *Applied Mathematics and Computation* 168.1 (Sept. 2005), pp. 722–737. ISSN: 0096-3003. DOI: `10.1016/j.amc.2005.01.142`.

[79] P. Shen. "Linearization method of global optimization for generalized geometric programming". In: *Applied Mathematics and Computation* 162.1 (Mar. 2005), pp. 353–370.

[80] Peiping Shen and Hongwei Jiao. "A new rectangle branch-and-pruning approach for generalized geometric programming". In: *Applied Mathematics and Computation* 183.2 (Dec. 2006), pp. 1027–1038. DOI: `10.1016/j.amc.2006.05.137`.

[81] Shao-Jian Qu, Ke-Cun Zhang, and Ying Ji. "A new global optimization algorithm for signomial geometric programming via Lagrangian relaxation". In: *Applied Mathematics and Computation* 184.2 (Jan. 2007), pp. 886–894. DOI: `10.1016/j.amc.2006.05.208`.

[82] P. Shen, Y. Ma, and Y. Chen. "A robust algorithm for generalized geometric programming". In: *Journal of Global Optimization* 41.4 (Aug. 2008), pp. 593–612. ISSN: 1573-2916. DOI: `10.1007/s10898-008-9283-0`.

[83] X. Hou, P. Shen, and Y. Chen. "A Global Optimization Algorithm for Signomial Geometric Programming Problem". In: *Abstract and Applied Analysis* 2014 (2014), pp. 1–12. DOI: `10.1155/2014/163263`.

[84] Gongxian Xu. "Global optimization of signomial geometric programming problems". In: *Euro. Journal of Operational Research* 233.3 (2014), pp. 500–510. DOI: `10.1016/j.ejor.2013.10.01`.

[85] D. H. Rountree and A. K. Rigler. "A penalty treatment of equality constraints in generalized geometric programming". In: *Journal of Optimization Theory and Applications* 38.2 (Oct. 1982), pp. 169–178. ISSN: 1573-2878. DOI: `10.1007/BF00934080`.

[86] James Yan. "Signomial programs with equality constraints: numerical solution and applications". PhD thesis. University of British Columbia, 1976. DOI: `10.14288/1.0065625`.

[87] Max M. J. Opgenoord, Brian S Cohen, and Warren W. Hoburg. *Comparison of Algorithms for Including Equality Constraints in Signomial Programming*. Tech. rep. ACDL TR-2017-1. MIT, Aug. 2017.

[88] D. Hilbert. "Ueber die Darstellung definiter Formen als Summe von Formenquadraten". In: *Mathematische Annalen* 32.3 (Sept. 1888), pp. 342–350.

[89] M. Kojima, S. Kim, and H. Waki. "Sparsity in sums of squares of polynomials". In: *Mathematical Programming* 103.1 (Dec. 2004), pp. 45–62.

[90] H. Waki et al. "Sums of Squares and Semidefinite Program Relaxations for Polynomial Optimization Problems with Structured Sparsity". In: *SIAM Journal on Optimization* 17.1 (Jan. 2006), pp. 218–242.

[91] J. Nie and J. Demmel. "Sparse SOS Relaxations for Minimizing Functions that are Summations of Small Polynomials". In: *SIAM Journal on Optimization* 19.4 (Jan. 2009), pp. 1534–1558.

[92] Amir Ali Ahmadi and Anirudha Majumdar. "DSOS and SDSOS Optimization: More Tractable Alternatives to Sum of Squares and Semidefinite Optimization". In: *SIAM Journal on Applied Algebra and Geometry* 3.2 (Jan. 2019). DOI: `10.1137/18m118935x`. URL: `https://doi.org/10.1137/18m118935x`.

[93] Philippe P. Pébay, J. Maurice Rojas, and David C. Thompson. "Optimization and NP_R-completeness of certain fewnomials". In: *Proceedings of the 2009 conference on Symbolic numeric computation*. ACM Press, 2009. DOI: 10.1145/1577190.1577212. URL: https://doi.org/10.1145/1577190.1577212.

[94] Mehdi Ghasemi and Murray Marshall. "Lower Bounds for Polynomials Using Geometric Programming". In: *SIAM Journal on Optimization* 22.2 (Jan. 2012), pp. 460–473. DOI: 10.1137/110836869.

[95] M. Ghasemi, J. B. Lasserre, and M. Marshall. "Lower bounds on the global minimum of a polynomial". In: *Computational Optimization and Applications* 57.2 (Sept. 2013), pp. 387–402. DOI: 10.1007/s10589-013-9596-x.

[96] Sadik Iliman and Timo de Wolff. "Amoebas, nonnegative polynomials and sums of squares supported on circuits". In: *Research in the Mathematical Sciences* 3.9 (Mar. 2016). DOI: 10.1186/s40687-016-0052-2.

[97] Sadik Iliman and Timo de Wolff. "Lower Bounds for Polynomials with Simplex Newton Polytopes Based on Geometric Programming". In: *SIAM Journal on Optimization* 26.2 (2016), pp. 1128–1146. DOI: 10.1137/140962425.

[98] Mareike Dressler, Sadik Iliman, and Timo de Wolff. "A Positivstellensatz for Sums of Nonnegative Circuit Polynomials". In: *SIAM Journal on Applied Algebra and Geometry* 1.1 (2017), pp. 536–555. DOI: 10.1137/16m1086303.

[99] Henning Seidler and Timo de Wolff. *An Experimental Comparison of SONC and SOS Certificates for Unconstrained Optimization*. 2018. eprint: arXiv:1808.08431.

[100] J. Wang. *Nonnegative polynomials and circuit polynomials*. 2018. eprint: arXiv:1804.09455.

[101] J. Wang. *On supports of sums of nonnegative circuit polynomials*. 2018. eprint: arXiv:1809.10608.

[102] O. Karaca et al. "The REPOP Toolbox: Tackling Polynomial Optimization Using Relative Entropy Relaxations". In: *IFAC-PapersOnLine* 50.1 (2017). 20th IFAC World Congress, pp. 11652–11657. ISSN: 2405-8963.

[103] G. M. Ziegler. *Lectures on Polytopes*. Springer, New York, 1995. DOI: 10.1007/978-1-4613-8431-1.

[104] Günter Ewald. *Combinatorial Convexity and Algebraic Geometry*. Springer New York, 1996. DOI: 10.1007/978-1-4612-4044-0. URL: https://doi.org/10.1007/978-1-4612-4044-0.

[105] L. Tunçel. "Generalization of Primal—Dual Interior-Point Methods to Convex Optimization Problems in Conic Form". In: *Foundations of Computational Mathematics* 1.3 (July 2001), pp. 229–254.

[106] A. Skajaa and Y. Ye. "A homogeneous interior-point algorithm for nonsymmetric convex conic optimization". In: *Mathematical Programming* 150.2 (May 2014), pp. 391–422.

[107] Santiago Akle Serrano. "Algorithms for unsymmetric cone optimization and an implementation for problems with the exponential cone". PhD thesis. Palo Alto, CA: Stanford University, 2015.

[108] D. Papp and S. Yıldız. *On "A Homogeneous Interior-Point Algorithm for Non-Symmetric Convex Conic Optimization"*. 2017. eprint: `arXiv:1712.00492`.

[109] Henning Seidler and Timo de Wolff. *POEM: Effective Methods in Polynomial Optimization, version 0.2.1.0(a)*. `http://www.iaa.tu-bs.de/AppliedAlgebra/POEM/index.html`. July 2019.

[110] M. Schweighofer. "An algorithmic approach to Schmüdgen's Positivstellensatz". In: *Journal of Pure and Applied Algebra* 166.3 (Jan. 2002), pp. 307–319.

[111] J.-L. Krivine. "Anneaux préordonnés". In: *Journal d'analyse mathématique* 12 (1964), p. 307–326.

[112] M. Marshall. *Positive polynomials and sums of squares*. Vol. 146. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2008, pp. xii+187.

[113] S. Diamond and S. Boyd. "CVXPY: A Python-Embedded Modeling Language for Convex Optimization". In: *Journal of Machine Learning Research* 17.83 (2016), pp. 1–5.

[114] A. Agrawal et al. "A Rewriting System for Convex Optimization Problems". In: *Journal of Control and Decision* 5.1 (2018), pp. 42–60.

[115] A. Domahidi, E. Chu, and S. Boyd. "ECOS: An SOCP solver for embedded systems". In: *European Control Conference (ECC)*. 2013, pp. 3071–3076.

[116] Hamza Fawzi. "Power and limitations of convex formulations via linear and semidefinite programming lifts". PhD thesis. Massachusettes Institute of Technology, 2016.

[117] Olga Kuryatnikova. "The many faces of positivity to approximate structured optimization problems". PhD thesis. Tilburg University, 2019.

[118] Olga Kuryatnikova, Juan C. Vera, and Luis F. Zuluaga. *The copositive way to obtain certificates of non-negativity over general semialgebraic sets*. 2019. eprint: `arXiv:1909.06689`.

[119] MOSEK ApS. *MOSEK 9.0.70(beta)*. 2019.

[120] Jonathan Borwein and Adrian Lewis. *Convex Analysis and Nonlinear Optimization*. Springer New York, 2006. DOI: `10.1007/978-0-387-31256-9`.

[121]   Didier Henrion, Jean-Bernard Lasserre, and Johan Löfberg. "GloptiPoly 3: moments, optimization and semidefinite programming". In: *Optimization Methods and Software* 24.4-5 (2009), pp. 761–779. DOI: `10.1080/10556780802699201`.

[122]   Monique Laurent. "Sums of Squares, Moment Matrices and Optimization Over Polynomials". In: *Emerging Applications of Algebraic Geometry*. Ed. by Mihai Putinar and Seth Sullivant. New York, NY: Springer New York, 2009, pp. 157–270. ISBN: 978-0-387-09686-5. DOI: `10.1007/978-0-387-09686-5_7`. URL: `https://doi.org/10.1007/978-0-387-09686-5_7`.

[123]   J. B. Lasserre. *An Introduction to Polynomial and Semi-Algebraic Optimization*. Cambridge University Press, 2015. DOI: `10.1017/cbo9781107447226`. URL: `https://doi.org/10.1017/cbo9781107447226`.

[124]   A. H. Wang et al. *A Positivstellensatz for Conditional SAGE*. 2020. eprint: `arXiv:2003.03731`.

[125]   C. Nowzari, V. M. Preciado, and G. J. Pappas. "Optimal Resource Allocation for Control of Networked Epidemic Models". In: *IEEE Transactions on Control of Network Systems* 4.2 (2017), pp. 159–169.

[126]   V. M. Preciado et al. "Optimal Resource Allocation for Network Protection Against Spreading Processes". In: *IEEE Transactions on Control of Network Systems* 1.1 (2014), pp. 99–108.

[127]   Charles N. Delzell. "Impossibility of extending Pólya's theorem to "forms" with arbitrary real exponents". In: *Journal of Pure and Applied Algebra* 212.12 (2008), pp. 2612–2622. ISSN: 0022-4049. DOI: `https://doi.org/10.1016/j.jpaa.2008.04.006`. URL: `http://www.sciencedirect.com/science/article/pii/S002240490800090X`.

[128]   J.B. Lasserre. *Moments, Positive Polynomials and their Applications*. Imperial College Press, London, 2010. DOI: `10.1142/p665`.

[129]   V. Powers and T. Wörmann. "An algorithm for sums of squares of real polynomials". In: *J. Pure and Applied Algebra* 127.1 (1998), pp. 99–104.

[130]   L. Katthän, H. Naumann, and T. Theobald. *A unified framework of SAGE and SONC polynomials and its duality theory*. To appear in *Math. Computation*. 2020.

[131]   G. Averkov. "Optimal size of linear matrix inequalities in semidefinite approaches to polynomial optimization". In: *SIAM J. Appl. Algebra and Geometry* 3.1 (2019), pp. 128–151. DOI: `10.1137/18M1201342`.

[132]   J. Wang and V. Magron. "A second order cone characterization for sums of nonnegative circuits". In: *Proc. International Symposium on Symbolic and Algebraic Computation (ISSAC)*. ACM, 2020, pp. 450–457.

[133] H. Naumann and T. Theobald. *The S-cone and a primal-dual view on second-order representability*. To appear in *Beiträge Algebra Geom.* 2020.

[134] D. Papp. *Duality of sum of nonnegative circuit polynomials and optimal SONC bounds*. Preprint, arXiv:1912.04718. 2019.

[135] Akshay Agrawal, Steven Diamond, and Stephen Boyd. "Disciplined geometric programming". In: *Optimization Letters* 13.5 (Mar. 2019), pp. 961–976. DOI: 10.1007/s11590-019-01422-z. URL: https://doi.org/10.1007/s11590-019-01422-z.

[136] W. Jarczyk and J. Matkowski. "On Mulholland's Inequality". In: *Proc. Amer. Math. Soc.* 130.11 (2002), pp. 3243–3247. ISSN: 00029939, 10886826. URL: http://www.jstor.org/stable/1194150.

[137] M. Özdemir, Ç. Yildiz, and M. Gürbüz. "A note on geometrically convex functions". In: *J. Inequalities and Applications* 2014.1 (2014), p. 180. DOI: 10.1186/1029-242x-2014-180. URL: https://doi.org/10.1186/1029-242x-2014-180.

[138] C. P. Niculescu. "Convexity according to the geometric mean". In: *Math. Inequalities & Applications* 3 (2 2000), pp. 155–167.

[139] I. M. Gelfand, M. M. Kapranov, and A. V. Zelevinsky. *Discriminants, Resultants, and Multidimensional Determinants*. Springer, 1994, pp. 271–296.

[140] B. Sturmfels. *Gröbner Bases and Convex Polytopes*. Amer. Math. Soc., 1996.

[141] F. Fillastre and I. Izmestiev. "Shapes of polyhedra, mixed volumes and hyperbolic geometry". In: *Mathematika* 63.1 (2017), pp. 124–183.

[142] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, Philadelphia, 2001. DOI: 10.1137/1.9780898718829. URL: https://doi.org/10.1137/1.9780898718829.

[143] J. B. Lasserre. "A New Look at Nonnegativity on Closed Sets and Polynomial Optimization". In: *SIAM Journal on Optimization* 21.3 (July 2011), pp. 864–885. DOI: 10.1137/100806990. URL: https://doi.org/10.1137/100806990.

[144] Nikolaos V. Sahinidis. "BARON: A general purpose global optimization software package". In: *Journal of Global Optimization* 8.2 (Mar. 1996), pp. 201–205. DOI: 10.1007/bf00138693. URL: https://doi.org/10.1007/bf00138693.

[145] Ruth Misener and Christodoulos A. Floudas. "ANTIGONE: Algorithms for coNTinuous / Integer Global Optimization of Nonlinear Equations". In: *Journal of Global Optimization* 59.2-3 (Mar. 2014), pp. 503–526. DOI: 10.1007/s10898-014-0166-2. URL: https://doi.org/10.1007/s10898-014-0166-2.

[146] Youdong Lin and Linus Schrage. "The global solver in the LINDO API". In: *Optimization Methods & Software* 24.4-5 (2009), pp. 657–668.

[147] Tobias Achterberg. "SCIP: solving constraint integer programs". In: *Mathematical Programming Computation* 1.1 (July 2009), pp. 1–41. ISSN: 1867-2957. DOI: 10.1007/s12532-008-0001-1. URL: https://doi.org/10.1007/s12532-008-0001-1.

[148] Gerald Gamrath et al. *The SCIP Optimization Suite 7.0*. Technical Report. Optimization Online, Mar. 2020. URL: http://www.optimization-online.org/DB_HTML/2020/03/7705.html.

[149] Jie Wang, Haokun Li, and Bican Xia. "A New Sparse SOS Decomposition Algorithm Based on Term Sparsity". In: *Proceedings of the 2019 on International Symposium on Symbolic and Algebraic Computation*. ISSAC '19. Beijing, China: Association for Computing Machinery, 2019, pp. 347–354. ISBN: 9781450360845. DOI: 10.1145/3326229.3326254. URL: https://doi.org/10.1145/3326229.3326254.

[150] Jie Wang, Victor Magron, and Jean-Bernard Lasserre. "TSSOS: A Moment-SOS Hierarchy That Exploits Term Sparsity". In: *SIAM Journal on Optimization* 31.1 (Jan. 2021), pp. 30–58. DOI: 10.1137/19m1307871. URL: https://doi.org/10.1137/19m1307871.

[151] P. J. C. Dickinson and J. Povh. "On an extension of Pólya's Positivstellensatz". In: *Journal of Global Optimization* 61.4 (May 2014), pp. 615–625. DOI: 10.1007/s10898-014-0196-9. URL: https://doi.org/10.1007/s10898-014-0196-9.

[152] Peter J. C. Dickinson and Janez Povh. "A new approximation hierarchy for polynomial conic optimization". In: *Computational Optimization and Applications* 73.1 (Jan. 2019), pp. 37–67. DOI: 10.1007/s10589-019-00066-0. URL: https://doi.org/10.1007/s10589-019-00066-0.

[153] E. de Klerk et al. "Bound-Constrained Polynomial Optimization Using Only Elementary Calculations". In: *Mathematics of Operations Research* 42.3 (Aug. 2017), pp. 834–853. DOI: 10.1287/moor.2016.0829. URL: https://doi.org/10.1287/moor.2016.0829.

[154] Gilles Cassier. "Probléme des moments sur un compact de Rn et décomposition de polynomes a plusieurs variables". In: *Journal of Functional analysis* 58.3 (1984), pp. 254–266.

[155] David Handelman. "Representing polynomials by positive linear functions on compact convex polyhedra". In: *Pacific Journal of Mathematics* 132.1 (1988), pp. 35–62.

[156] Etienne De Klerk, Monique Laurent, and Zhao Sun. "Convergence analysis for Lasserre's measure-based hierarchy of upper bounds for polynomial optimization". In: *Mathematical Programming* 162.1-2 (2017), pp. 363–392.

[157] Etienne De Klerk, Roxana Hess, and Monique Laurent. "Improved convergence rates for Lasserre-type hierarchies of upper bounds for box-constrained polynomial optimization". In: *SIAM Journal on Optimization* 27.1 (2017), pp. 347–367.

[158] Etienne de Klerk and Monique Laurent. "Convergence analysis of a Lasserre hierarchy of upper bounds for polynomial minimization on the sphere". In: *Mathematical Programming* (2020), pp. 1–21.

[159] Etienne de Klerk and Monique Laurent. "Worst-case examples for Lasserre's measure–based hierarchy for polynomial optimization on the hypercube". In: *Mathematics of Operations Research* 45.1 (2020), pp. 86–98.

[160] Lucas Slot and Monique Laurent. "Improved convergence analysis of Lasserre's measure-based upper bounds for polynomial minimization on compact sets". In: *Mathematical Programming* (2020), pp. 1–41.

[161] M. K. de Carli Silva and L. Tunçel. "A Notion of Total Dual Integrality for Convex, Semidefinite, and Extended Formulations". In: *SIAM Journal on Discrete Mathematics* 34.1 (Jan. 2020), pp. 470–496. DOI: 10.1137/18m1169710. URL: https://doi.org/10.1137/18m1169710.

[162] M. J. Rijckaert and X. M. Martens. "Comparison of generalized geometric programming algorithms". In: *Journal of Optimization Theory and Applications* 26.2 (Oct. 1978), pp. 205–242. ISSN: 1573-2878. DOI: 10.1007/BF00933404.

[163] Jerome Bracken and Garth P McCormick. *Selected applications of nonlinear programming*. 1968.

[164] Gary E. Blau and Douglass J. Wilde. "Generalized polynomial programming". In: *The Canadian Journal of Chemical Engineering* 47.4 (Aug. 1969), pp. 317–326. DOI: 10.1002/cjce.5450470401. URL: https://doi.org/10.1002/cjce.5450470401.

[165] M. J. D. Powell. "A Direct Search Optimization Method That Models the Objective and Constraint Functions by Linear Interpolation". In: *Advances in Optimization and Numerical Analysis*. Dordrecht: Springer Netherlands, 1994, pp. 51–67. ISBN: 978-94-015-8330-5.

[166] J. B. Lasserre and E. S. Zeron. "Solving a class of multivariate integration problems via Laplace techniques". In: *Applicationes Mathematicae* 28.4 (2001), pp. 391–405. DOI: 10.4064/am28-4-2. URL: https://doi.org/10.4064/am28-4-2.

[167] A. I. Barvinok. "Computing the volume, counting integral points, and exponential sums". In: *Discrete Comput. Geom.* 10.2 (1993), pp. 123–141.

[168] Gregory V. Bard. "Some Basic Facts about Linear Algebra over GF(2)". In: *Algebraic Cryptanalysis*. Springer US, 2009, pp. 81–88. DOI: 10.1007/978-0-387-88757-9_6.

[169] Jan Verschelde. "Algorithm 795: PHCpack: A General-purpose Solver for Polynomial Systems by Homotopy Continuation". In: *ACM Trans. Math. Softw.* 25.2 (June 1999), pp. 251–276. ISSN: 0098-3500. DOI: `10.1145/317275.317286`.

[170] Shashwati Ray and P. S. V. Nataraj. "An efficient algorithm for range computation of polynomials using the Bernstein form". In: *Journal of Global Optimization* 45.3 (Dec. 2008), pp. 403–426. DOI: `10.1007/s10898-008-9382-y`.

[171] Jean B. Lasserre, Kim-Chuan Toh, and Shouguang Yang. "A bounded degree SOS hierarchy for polynomial optimization". In: *EURO Journal on Computational Optimization* 5.1 (Mar. 2017), pp. 87–117. ISSN: 2192-4414. DOI: `10.1007/s13675-015-0050-y`.

[172] Tillmann Weisser, Jean B. Lasserre, and Kim-Chuan Toh. "Sparse-BSOS: a bounded degree SOS hierarchy for large scale polynomial optimization with sparsity". In: *Mathematical Programming Computation* 10.1 (Mar. 2018), pp. 1–32. ISSN: 1867-2957. DOI: `10.1007/s12532-017-0121-6`.

[173] Antonis Papachristodoulou et al. *SOSTOOLS Version 3.00 Sum of Squares Optimization Toolbox for MATLAB*. 2013. eprint: `arXiv:1310.4716`.

[174] Edward Burnell, Nicole B Damen, and Warren Hoburg. "GPkit: A Human-Centered Approach to Convex Optimization in Engineering Design". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020.

[175] Victoria Powers and Bruce Reznick. "Polynomials that are positive on an interval". In: *Transactions of the American Mathematical Society* 352.10 (Oct. 2000), pp. 4677–4692. DOI: `10.1090/s0002-9947-00-02595-2`.

[176] S. Surjanovic and D. Bingham. *Virtual Library of Simulation Experiments: Test Functions and Datasets*. Retrieved April 18, 2019, from `http://www.sfu.ca/~ssurjano`.

[177] Lieven Vandenberghe. *The CVXOPT linear and quadratic cone program solvers*. `http://www.seas.ucla.edu/~vandenbe/publications/coneprog.pdf`. Mar. 2010.

[178] Chris Coey, Lea Kapelevich, and Juan Pablo Vielma. *Towards practical generic conic optimization*. 2021. arXiv: `2005.01136 [math.OC]`.