UNIQUE POWERS-OF-FORMS DECOMPOSITIONS FROM SIMPLE GRAM SPECTRAHEDRA

ALEXANDER TAVEIRA BLOMENHOFER

Abstract. We consider simultaneous Waring decompositions: Given forms f_d of degrees kd, $(d = 2, 3)$, which admit a representation as d-th power sums of k-forms $q_1, \ldots, q_m$, when is it possible to reconstruct the addends $q_1, \ldots, q_m$ from the power sums f_d? Such powers-of-forms decompositions model the moment problem for mixtures of centered Gaussians. The novel approach of this paper is to use semidefinite programming in order to perform a reduction to tensor decomposition. The proposed method works on typical parameter sets at least as long as $m \leq n-1$, where m is the rank of the decomposition and n is the number of variables. While provably not tight, this analysis still gives the currently best known rank threshold for decomposing third order powers-of-forms, improving on previous work [30], which required $\Omega(1) \leq m \leq \mathcal{O}(\sqrt{n})$ and, more recently, Bafna, Hsieh, Kothari and Xu [6], which can go up to $m = \mathcal{O}(\frac{n}{\log(n)^2})$. Our algorithm can produce proofs of uniqueness for specific decompositions. A numerical study is conducted on Gaussian random trace-free quadratics, giving evidence that the success probability converges to 1 in an average case setting, as long as $m = n$ and $n \to \infty$. Some evidence is given that the algorithm also succeeds on instances of rank $m = \Theta(n^2)$.

1. Introduction

Waring decompositions for polynomials are a highly studied problem with a wide range of applications in sciences and statistics, including phylogenetics [39], cryogenic electron microscopy [7], Gaussian mixtures ([30],[29], see Section 5) and many more [39]. They serve as a fundamental model in the theory of arithmetic circuits [29] and occur as an important algorithmic primitive for various machine learning problems [8],[46]. Formally, for a fixed form, i.e., a homogeneous polynomial $f_d \in K[X]_{dk}$ of degree $dk \in \mathbb{N}$, the representation

$$f_d = \sum_{i=1}^{m} \lambda_i q_i^d, \qquad (q_1, \ldots, q_m \in K[X]_k,\ \lambda \in K^m) \tag{1.1}$$

is a *k-Waring decomposition* of *rank* m of f_d over the field K. For $f_d \in K[X]$, the minimum m such that f_d has a k-Waring decomposition of rank m is called the *k-Waring rank* of f_d over the field K. Classically, the main focus of attention used to be the case where $K = \mathbb{C}$ is the field of complex numbers, and power sums of *linear* polynomials were considered. The latter corresponds to $k = 1$. Note that over the complex field, the weights λ_i are redundant, and thus omitted. A long series of work, started more than a century ago, e.g., by Sylvester [66], Hilbert [34] and Terracini [68], lead via the celebrated Alexander-Hirschowitz theorem [35] and results by Chiantini-Ottaviani-Vannieuwenhoven [16],[17],[18], Galuppi-Mella

Date: May 12, 2023.
2020 *Mathematics Subject Classification.* primary: 14Q30, 15A69. secondary: 62H12, 68Q25.
Key words and phrases. Waring decomposition, SDP, Gaussian mixtures, subspace learning.
Centrum Wiskunde & Informatica, Amsterdam.

[28] (and many others) to a complete classification when *generic* 1-Waring decompositions are unique for the forms they describe. This property is called *generic identifiability*.

A second line of thoughts (e.g., [31],[42],[3],[22]) developed from the side of applications and was continued by theoretical computer scientists. Among many other things, it produced *algorithmic uniqueness theorems* for 1-Waring decompositions. From one perspective, these can be seen as "efficient" algorithms that recover the representation $f_d = \ell_1^d + \ldots + \ell_m^d$ from the d-th power sum f_d as input, under some restrictive assumptions on the rank and some nondegeneracy assumptions on the parameters of the representation, cf. Theorem 2.3 and [12, Theorem 2.4.8]. From another perspective, these results provide a proof of uniqueness for the minimum rank decomposition, whenever certain explicit conditions are met. Therefore, they are a tool to produce rank lower bounds for explicit families of polynomials.

The work on higher degree Waring decompositions has been pioneered by Reznick [56],[57],[63] (with Tokcan), [58],[59],[60],[61],[62]. In recent years, geometers have been trying to understand uniqueness and generic rank of Waring decompositions also for higher values of k [26]. A conjecture due to Ottaviani [45, Conjecture 1.2] states that the generic rank of k-Waring decompositions behaves as expected from counting parameters, if $d \geq 3$. In [13], the author showed in joint work with Casarotti, Michalek and Oneto, that for "most" subgeneric ranks m, Waring decompositions are unique, based on work of Nenashev [49] and Casarotti-Mella [14]. The results imply in particular bounds on the generic rank for k-Waring decompositions. Casarotti and Postinghel [15] then studied a different asymptotic setting where not the number of variables n but rather the degree k is assumed to be large. Simultaneous Waring decompositions for vectors of forms of (possibly) different degrees have been studied geometrically e.g. by Angelini, Galuppi, Mella and Ottaviani [5]. From the computational perspective, the work of Garg, Kayal and Saha [29] and Bafna, Hsieh, Kothari, Xu [6] examined polynomial-time recovery procedures for some variants of powers-of-forms decomposition.

This paper aims to generalize the second line of work, concerning algorithmic uniqueness theorems, to higher values of k, although a slightly different setting is considered: The focus is on *real* decompositions, degree $k \geq 2$, and simultaneous power sum decomposition in various degrees. Formally, we provide an algorithm and a uniqueness theorem (cf. Algorithm 1 and Theorem 3.1) for third order powers-of-forms (POF) decompositions, which have the following basic template:

$$\begin{aligned} (\mathrm{PoF})_{f,m,k} \qquad \text{given} \quad & f_0, f_1, f_2, f_3 \text{ of degrees } 0, k, 2k, 3k, \\ \text{find} \quad & q_1, \ldots, q_m \in \mathbb{R}[X]_k, \\ & \lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}, \\ \text{s. t.} \quad & f_d = \sum_{i=1}^{m} \lambda_i q_i^d, \qquad d = 0, 1, 2, 3. \end{aligned} \tag{1.2}$$

Let us call f_0, f_1, f_2, f_3 the *power sums*, $q_1, \ldots, q_m$ the *addends* and $\lambda_1, \ldots, \lambda_m$ the (nonnegative) *weights*. Throughout, $X = (X_1, \ldots, X_n)$ are polynomial variables and $n, m, k \in \mathbb{N}$ are positive integers. For the scope of this paper, we limit our attention to real POF decompositions. The recovery task is not necessarily well-posed, since a given form f might have various decompositions. However, if m is not too large, e.g., if $m = \mathcal{O}(n^{(d-1)k})$, then a general form of k-Waring rank m has a unique decomposition, cf. [13].

The special case $k = 1$ relates to a plethora of important problems, e.g.: atom reconstruction of finitely supported measures, mixtures of Gaussians with *identical* covariance matrices and to symmetric tensor decomposition [12]. The case $k = 2$ of

quadratic forms has to do with mixtures of *centered* Gaussians. Therefore, it has Machine Learning applications, e.g., for learning a union of subspaces. These connections are explained in Section 5. For mixtures of centered Gaussians, third order powers-of-quadratics decomposition yields the first case where nontrivial recovery results are achievable. This case is a special focus of the present paper, although our algorithm works for all values of k.

1.1. **Overview of Contributions and main results.** This paper proposes and analyzes an algorithm to recover third order POF decompositions, as stated in (1.2). A proof-of-concept implementation of the algorithm in `Julia` can be found on GitHub, see [67]. Algorithm 1 is based on semidefinite programming and is "efficient" in the sense that, aside from its calls to the SDP solver, which is treated as a blackbox, it only uses basic linear-algebraic operations on polynomially-sized quantities constructed from the input. In other words, one could say that it is efficient up to the automatizability of semidefinite programming, which is still not completely understood, cf. [50]. Any numerical troubles, such as condition, are also ignored. E.g., for the sake of readability, we will write "$\lambda_1 > 0$" rather than requiring λ_1 to be sufficiently bounded away from zero.

When it succeeds, the algorithm will also produce a proof of uniqueness of the decomposition. Therefore, it implies an algorithmic uniqueness result, Theorem 3.1, which is at the core of this paper. The conditions of Theorem 3.1 can be explicitly described in terms of just the second power sum f_2 of degree $2k$ and checked *before* computing the decomposition. One basic tool is to associate a subspace of degree-k forms to the second order power sum f_2, which will be called the *Sum of Squares support* of f_2. It consists of all polynomials contributing to *some* Sum-of-Squares decomposition of f_2:

$$\operatorname{sosupp} f_2 = \{p \in \mathbb{R}[X]_k \mid \exists \lambda \in \mathbb{R}_{>0} \colon f_2 - \lambda p^2 \text{ is a sum of squares}\}. \tag{1.3}$$

The Sum of Squares support will be explained in detail in Section 2. For now, it was just introduced in order to state the main result.

Theorem 1.1 (Cf. Theorem 3.1)**.** *Let $k \in \mathbb{N}$ and let f_2, f_3 be forms of degree $2k$ and $3k$, respectively. Denote $U := \operatorname{sosupp} f_2$ and $N := \dim U$. Assume that the space of threefold products $\{u \cdot v \cdot w \mid u, v, w \in U\}$ of the polynomials in U has dimension $\binom{N+2}{3}$.*[1] *Then f_2 and f_3 have at most one joint POF decomposition*

$$f_d = \sum_{i=1}^{m} \lambda_i q_i^d, \qquad d = 2, 3 \tag{1.4}$$

with positive weights $\lambda_1, \ldots, \lambda_m > 0$, $m \in \mathbb{N}$ and linearly independent $q_1, \ldots, q_m$. Furthermore, if such a decomposition exists, then Algorithm 1 computes it efficiently and it is the unique minimum rank POF decomposition of (f_2, f_3).

Unlike Waring decompositions of order 3 or higher, Sum-of-Squares representations are highly non-unique. Indeed, consider a Sum-of-Squares representation $f = p_1^2 + \ldots + p_N^2$ of some form f and any orthogonal matrix $A \in \mathbb{R}^{N \times N}$. Write $p := (p_1, \ldots, p_N)$. Then clearly also the entries of $s := Ap$ give a Sum-of-Squares representation of f, since

$$s_1^2 + \ldots + s_N^2 = s^T s = p^T A^T A p = p^T p = p_1^2 + \ldots + p_N^2 = f. \tag{1.5}$$

Nevertheless, we say that a form $f \in \mathbb{R}[X]_{2k}$ is *uniquely Sum-of-Squares representable*, if there is only one Sum-of-Squares representation modulo orthogonal transformations.

[1] In other words, there are no algebraic relations between linearly independent elements of U of degree at most 3.

Section 3.1 derives and explains the idea behind Algorithm 1 and states the main result, while Section 4 offers a detailed discussion of the conditions. Here, we just highlight a few corollaries for instances constructed from *general* addends $q_1, \dots, q_m$. One important quantity will be the number $\beta_k(n)$ introduced below.

Definition 1.2. *Let $n \in \mathbb{N}$. Then we denote by $\beta_k(n)$ the maximum number m of k-forms $q_1, \dots, q_m$ that satisfy one of the following equivalent conditions*

(1) There are no algebraic relations of $q_1, \dots, q_m$ of degree at most 3.
(2) There are no homogeneous algebraic relations of $q_1, \dots, q_m$ of degree 3.
(3) $\dim \mathbb{R}[q_1, \dots, q_m]_{3k} = \binom{m+2}{3}$.

It holds that $\beta_k(n) \geq n+1$ for all $k \geq 2, n \geq 2$ and, e.g., that $\beta_2(n) = \Theta(n^2)$, see Section 4.1. For power sums constructed from at most $m \leq \beta_k(n)$ *general* addends, the conditions of Theorem 1.1 are satisfied, if f_2 is uniquely Sum-of-Squares representable. See Section 4.1 and, in particular, Proposition 4.4 for a detailed discussion. This yields the following simplification.

Corollary 1.3. *For any $n, m, k \in \mathbb{N}$ with $m \leq \beta_k(n)$, there is an efficient algorithm for the following problem: If $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{>0}$ and $q_1, \dots, q_m \in \mathbb{R}[X]_k$ are general forms such that $f_2 := \sum_{i=1}^m \lambda_i q_i^2$ is uniquely Sum-of-Squares representable, compute the set $\{(q_1, \lambda_1), \dots, (q_m, \lambda_m)\}$ from inputs f_2 and $f_3 := \lambda_1 q_1^3 + \dots + \lambda_m q_m^3$.*

For $m < n$, it is in addition possible to give geometric criteria.

Corollary 1.4. *For any $n, m \in \mathbb{N}$ with $m \leq n-2$, there is an efficient algorithm for the following problem: If $q_1, \dots, q_m \in \mathbb{R}[X]_k$ are general forms such that their real variety $V_{\mathbb{R}}(q_1, \dots, q_m)$ contains a nonzero point and $\lambda_1, \dots, \lambda_m$ are positive, compute the set $\{(q_1, \lambda_1), \dots, (q_m, \lambda_m)\}$ from inputs $f_2 := \lambda_1 q_1^2 + \dots + \lambda_m q_m^2$ and $f_3 := \lambda_1 q_1^3 + \dots + \lambda_m q_m^3$.*

The case $m = n-1$ is special:

Corollary 1.5. *For any $n \in \mathbb{N}$, there is an efficient algorithm for the following problem: If $q_1, \dots, q_{n-1} \in \mathbb{R}[X]_k$ are general forms such that all the finitely many lines of the variety $V(q_1, \dots, q_{n-1})$ are real and $\lambda_1, \dots, \lambda_m$ are positive, compute the set $\{(q_1, \lambda_1), \dots, (q_m, \lambda_m)\}$ from inputs $f_2 := \lambda_1 q_1^2 + \dots + \lambda_m q_m^2$ and $f_3 := \lambda_1 q_1^3 + \dots + \lambda_m q_m^3$.*

In Appendix C, it is proven that the conditions of both Corollary 1.4 and Corollary 1.5 are satisfied for *typical* choices of $q_1, \dots, q_m$, i.e. on a Euclidean open subset of $\mathbb{R}[X]_k^m$. However, note that the geometric arguments fail for $m \gg n$, whereas one has evidence to believe that Algorithm 1 can also decompose some instances of quadratic rank $m = \Theta(n^2)$, cf. Conjecture 4.10. Therefore it is not clear whether Corollary 1.3 extends beyond $m = n-1$, but numerical evidence from Section 4.2 strongly suggests so. Unique Sum-of-Squares representability appears to be in general not well-understood and Corollary 1.3 gives further motivation to understand it better. The recovery result from Theorem 3.1 has consequences for certain Machine Learning problems, two of which we highlight in the following section.

1.2. **Learning parameters of centered Gaussian mixtures.** The parameter estimation problem for Gaussian mixtures has a rich history, dating back to Pearson [51]. It has now been studied over more than a century in all kinds of flavours e.g. from the perspective of computer science ([20],[64],[21],[47],[38],[37],[4],[54],[44]), algebraic geometry ([1],[2]), moment problems ([19],[23]) and applications ([55],[52]).

A mixture of m *centered* Gaussians has a degree-$2d$ moment form (cf. Section 5.1) proportional to

$$\sum_{i=1}^{m} \lambda_i q_i^d, \tag{1.6}$$

where $\lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}$ are the mixing weights (summing up to 1) and $q_1, \ldots, q_m$ are positive (semi)definite quadratic forms $q_i = X^T \Sigma_i X$, where Σ_i is the covariance matrix of the i-th centered Gaussian. Thus, there is a straightforward connection between the parameter estimation problem for mixtures of centered Gaussians from their moments on one side and decompositions as powers of quadratic forms on the other side. At first glance, Theorem 1.1 does not fare well together with the setting of Gaussian mixtures, since if one of the forms $q_1, \ldots, q_m$ is positive definite, then $\sum_{i=1}^m \lambda_i q_i^2$ will never be uniquely Sum-of-Squares representable. With some slight adaptations, it is possible to prove a recovery result for typical instances of Gaussian mixtures. This is done in Section 5 and highlighted here:

Corollary 1.6. *For any $n \in \mathbb{N}$, $m \in \{1, \ldots, n-1\}$, there is a Euclidean open subset $\mathcal{U}$ of $\mathbb{R}[X_1, \ldots, X_n]_2^m$ and an efficient algorithm for the following problem: If Y is a mixture of m centered Gaussian random variables with general positive definite covariance forms $(q_1, \ldots, q_m) \in \mathcal{U}$ and positive mixing weights $\lambda_1, \ldots, \lambda_m$, compute the set of parameters $\{(q_1, \lambda_1), \ldots, (q_m, \lambda_m)\}$ from the moments $\mathcal{M}_{\leq 6}(Y)$ of Y of degree at most 6.*

1.3. **Learning unions of subspaces.** A special type of Gaussian mixture distributions can be used as a model for subspace learning. Here, data is assumed to be normally distributed on either of the r-dimensional subspaces $U_1, \ldots, U_m$ and the task is to find bases for the subspaces $U_1, \ldots, U_m$ from samples of the mixture distribution as input. The main difference to a general Gaussian mixture instance from above is that the forms $q_1, \ldots, q_m$ corresponding to the subspaces $U_1, \ldots, U_m$ will not have full rank.

We highlight this special application, since it is a case where one needs uniqueness not for *general* forms, but for forms that are general *within the class of fixed-rank quadratic forms*. We are not aware of any decomposition result applicable for this case, since the previous work [6], [30] is based on a probabilistic analysis and thus implicitly assumes full-rank quadratics.

Corollary 1.7. *For any $n, r \in \mathbb{N}_{\geq 3}$, $m \leq n-1$, there is a Euclidean open subset $\mathcal{U}$ of the problem parameters[2] and an efficient algorithm for the following problem: If $Y_1, \ldots, Y_m$ are normally distributed random variables on r-dimensional subspaces $U_1, \ldots, U_m$ and $\lambda \in \mathbb{R}^m_{>0}$ with $\sum_{i=1}^m \lambda_i = 1$, compute bases for the subspaces $U_1, \ldots, U_m$ from the moments of $\lambda_1 Y_1 \oplus \ldots \oplus \lambda_m Y_m$ of degree at most 6.*

1.4. **Relevance of results.** The aim of this work is to get tighter, algorithmic rank lower bounds for third-order powers-of-forms decomposition. In a typical real case, with Corollary 1.5, we improve the rank threshold for efficient recovery from $m \leq \mathcal{O}(\frac{n}{\log(n)^2})$ (due to [6]) or $\Omega(1) \leq m \leq \mathcal{O}(\sqrt{n})$ (due to [30]) to $1 \leq m \leq n-1$. Compared to the previous results, this gives an improvement of the asymptotic order *and* the constant factors, with a much simpler proof. In addition, Algorithm 1 implicitly produces a proof of uniqueness of the minimum rank decomposition for any concrete instance where it succeeds.

Analysis beyond the case $m = n-1$ is more difficult, since one may not rely on geometric arguments any more. However, there is significant reason to hope that

[2]The parameters are the subspaces together with the means and covariances of the Gaussians.

Algorithm 1 can decompose instances of rank $m = \Theta(n^2)$. Indeed, $m = \Theta(n^2)$ general quadratics $q_1, \ldots, q_m$ do not satisfy any algebraic relations of degree 3, which is implicitly shown by Bafna, Hsieh, Kothari and Xu [6, Section 6.4]. The big open question is whether there are also instances of $m = \Theta(n^2)$ quadratics that are uniquely Sum-of-Squares representable, and if these sets have nonempty intersection. Numerical findings suggest this is the case, cf. Conjecture 4.10. The threshold $m \in \mathcal{O}(n)$ is thus likely not an actual algorithmic boundary, see Section 4.2.

Acknowledgements. I wish to thank Pravesh Kothari, who encouraged me to work on powers-of-forms decompositions and pointed me towards the work of Ankit Garg [29]. I also wish to thank Monique Laurent for the rich feedback provided to this article, in particular for catching several mistakes. I further wish to thank Greg Blekhermann and João Gouveia for the suggestion to study trace-free quadratic forms, Julian Vill and Claus Scheiderer for sharing their expertise on Gram Spectrahedra and to Simon Telen for being a good listener. Part of this work was completed while the author was supported by the Dutch Scientific Council (NWO) grant OCENW.GROOT.2019.015 (OPTIMAL).

Disclosure. The main ideas of this paper were published first as part of my doctoral thesis [12] at Universität Konstanz. Some formulations might therefore overlap. However, this paper significantly elaborates on the ideas that were present in [12].

2. Preliminaries

Notation. Let us write $\mathbb{N} = \{1, 2, 3, \ldots\}$ for the set of natural numbers and $\mathbb{N}_0$ for $\mathbb{N} \cup \{0\}$. This paper concerns POF decompositions over the real field $\mathbb{R}$, but we might occasionally mention some results that hold over the complex numbers $\mathbb{C}$. For $K \in \{\mathbb{R}, \mathbb{C}\}$, we endow any finite dimensional K-vector space U with the K-Zariski topology. The varieties considered in this paper are closed affine or projective varieties. Closed affine varieties are subsets of U that can be written as the feasible set $V(q_1, \ldots, q_m)$ of a system of polynomial equations

$$q_1(x) = 0, \ldots, q_m(x) = 0, \quad (x \in U). \tag{2.1}$$

The space of linear functionals from U to $\mathbb{R}$ is denoted $U^\vee$ and called the *dual space* of U. Algebraic unknowns will be denoted by capital letters. In particular, for $U = K^n$, it is by default assumed that the unknowns are $X = (X_1, \ldots, X_n)$ and the polynomial ring is denoted $K[X]$. Note that $p \in K[X]$ denotes a polynomial, whereas $p(x)$ denotes the evaluation of p in some point $x \in K^n$. As one exception, when talking about algebraic relations of some polynomials $q_1, \ldots, q_m \in K[X]$, let us denote their *ideal of relations*

$$I_{\mathrm{rel}}(q_1, \ldots, q_m) = \{f \in K[Y] \mid f(q_1, \ldots, q_m) = 0\} \tag{2.2}$$

in some separate set of unknowns $Y = (Y_1, \ldots, Y_m)$, to avoid confusion. For some graded K-algebra R, R_k denotes the k-th graded component of R and $R_{\leq k} := R_0 \oplus \ldots \oplus R_k$ denotes the part of grade at most k. Quotients of polynomial rings by homogeneous ideals will naturally inherit the grading by the degree. However, for a subalgebra $K[q_1, \ldots, q_m] \subseteq K[X_1, \ldots, X_n]$ generated by some k-forms $q_1, \ldots, q_m$, we will often deviate from the canonical grading by the degree and instead grade $K[q_1, \ldots, q_m]$ by $\frac{1}{k}$ times the degree, for technical reasons.

The reader is assumed to have some basic familiarity with convex geometry (e.g., the notions of convex cones, faces, relative interior, conic duality) and with algebraic geometry (e.g., Bertini's theorem). For the background knowledge, cf. the books of

Barvinok [9] for convex geometry and Hartshorne [32] for algebraic geometry. For a convex cone $C \subseteq U$ in some $\mathbb{R}$-vector space U, the *dual cone* of C is denoted as

$$C^* := \{L \in U^\vee \mid \forall u \in U\colon L(u) \geq 0\} \subseteq U^\vee. \tag{2.3}$$

In the special case where C is even a subspace of U, it holds

$$C^* = \{L \in U^\vee \mid \forall u \in U\colon L(u) = 0\}. \tag{2.4}$$

Thus, C^* is then a subspace of $U^\vee$, which is called the *conormal space* of C. It is commonly denoted $C^\perp$ rather than C^*. If R is a commutative, graded $\mathbb{R}$-algebra with graded components $R_0, R_1, R_2, \ldots$, then for $k \in \mathbb{N}_0$, we denote by

$$\Sigma_{R,2k} = \left\{ f \in R_{2k} \mid \exists N \in \mathbb{N}_0, p_1, \ldots, p_N \in R_k\colon f = \sum_{i=1}^N p_i^2 \right\}. \tag{2.5}$$

the *homogeneous Sums-of-Squares cone of* R in degree $2k$. If $R = \mathbb{R}[X]$ is the polynomial ring, we simply write Σ_{2k}, suppressing the dependency on the variables. For a homogeneous ideal $I \subseteq R$, we denote by I_k the degree-k component of I, i.e., $I_k = I \cap R_k$.

2.1. **Gram Spectrahedra.** A form $f \in \mathbb{R}[X]$ is a *sum of squares* if there exist $N \in \mathbb{N}_0$ and forms $q_1, \ldots, q_N \in \mathbb{R}[X]$ such that

$$f = \sum_{i=1}^N q_i^2. \tag{2.6}$$

The right hand side of (2.6) is called a *Sum-of-Squares representation*. For any orthogonal transformation $A \in \mathbb{R}^{N\times N}$, both $q = (q_1, \ldots, q_N)^T$ and $A(q_1, \ldots, q_N)^T$ represent the same polynomial f. Let us denote by $[X]_k = (X^\alpha)_{|\alpha|=k}$ the vector of monomials of degree k. Then, any polynomial $p \in \mathbb{R}[X]_k$ can be written as $p = c_p^T[X]_k$ for some real *coefficient vector* $c_p = (c_{p,\alpha})_{|\alpha|=k}$. This allows to write Sum-of-Squares representations such as (2.6) via *Gram matrices*

$$f = [X]_k^T \left(\sum_{i=1}^N c_{q_i} c_{q_i}^T \right) [X]_k = [X]_k^T G(q) [X]_k. \tag{2.7}$$

Here, we denote $G(q) := \sum_{i=1}^N c_{q_i} c_{q_i}^T$ for the positive semidefinite (psd) Gram matrix of the Sum-of-Squares representation $f = q^T q$. Let us write $G \succeq 0$ to denote that some (symmetric) matrix G is psd. It turns out that any matrix representation $f = [X]_k^T G [X]_k$, where $G \succeq 0$, corresponds to a class of Sum-of-Squares representations modulo orthogonal transformations. The convex set

$$\mathrm{Gram}(f) := \{G \succeq 0 \mid [X]_k^T G [X]_k = f\} \tag{2.8}$$

is called the *Gram spectrahedron* of f. Let us collect some basic properties.

Proposition 2.1. *Let $k \in \mathbb{N}$, $f \in \Sigma_{2k}$.*

(a) Every face F of $\mathrm{Gram}(f)$ *has an associated subspace U_F such that*

$$F = \{G \in \mathrm{Gram}(f) \mid \operatorname{im} G \subseteq U_F\}$$

and such that equality $\operatorname{im} G = U_F$ *holds for* all *points in the relative interior of F. We interpret U_F as a subspace of $\mathbb{R}[X]_k$, by sending $c \in U_F$ to $[X]_k^T c$.*

(b) A relative interior point of F corresponds to a class of Sum-of-Squares representations of f (modulo orthogonal transformations) of length $\dim U_F$.

(c) A linear subspace U of $\mathbb{R}[X]_k$ is called facial *for* $\mathrm{Gram}(f)$*, if there exists some G in* $\mathrm{Gram}(f)$ *such that* $\operatorname{im} G = U$.

(d) If $F' \subsetneq F$ is a proper subface, then $\dim U_{F'} < \dim U_F$.

(e) The set

$$\operatorname{sosupp} f = \{p \in \mathbb{R}[X]_k \mid \exists \lambda \in \mathbb{R}_{>0} \colon f - \lambda p^2 \text{ is a sum of squares}\}$$

of all polynomials contributing to some Sum-of-Squares decomposition of f is a subspace of $\mathbb{R}[X]_k$.

(f) The sum of facial subspaces is facial. In particular, there exists a largest facial subspace $U_{\operatorname{Gram}(f)}$ of $\operatorname{Gram}(f)$ and this subspace equals $\operatorname{sosupp} f$.

Proof. Cf. the work of Ramana and Goldman [53] on the facial structure of (arbitrary) spectrahedra. This formulation loosely follows Scheiderer [65, Section 2]. □

Recall that for a point x in some convex set C, the *supporting face* $\operatorname{suppf} x$ of x is defined to be the minimal face of C containing x. For a face F of C it holds $x \in \operatorname{relint} F$, if and only if F is the supporting face of x.

2.2. **Powers-of-forms decomposition.** Throughout, we will consider third order powers-of-forms decomposition, as introduced in (1.2). Let us restate the standard setting of third order POF decomposition used in this paper:

$$\begin{aligned} (\mathrm{PoF})_{f,m,k} \qquad \text{given} \quad & f_0, f_1, f_2, f_3 \text{ of degrees } 0, k, 2k, 3k, \\ \text{find} \quad & q_1, \dots, q_m \in \mathbb{R}[X]_k, \\ & \lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}, \\ \text{s. t.} \quad & f_d = \sum_{i=1}^{m} \lambda_i q_i^d, \qquad d = 0, 1, 2, 3. \end{aligned} \tag{2.9}$$

Note that in order to recover both the addends and the weights, it is necessary to use power sums of at least two different orders. Our main results, Theorem 3.1 and Algorithm 1, are "minimal" in the sense that they only make use of the power sums f_2 and f_3. In some applications, it is canonical and useful to have f_1 as well. This is explained in Section 5.

Definition 2.2. *Let $k, d \in \mathbb{N}$ and $f \in \mathbb{R}[X]_{dk}$. There exists a smallest number $m \in \mathbb{N}$ such that f has a k-*Waring decomposition *of* rank *m, i.e., a decomposition*

$$f = \sum_{i=1}^{m} \sigma_i q_i^d \tag{2.10}$$

of f as a signed sum of m d-th powers of k-forms $q_1, \dots, q_m$, with signs *$\sigma_1, \dots, \sigma_m \in \{\pm 1\}$. This m is called the (real) k-*Waring rank *of f. For odd d, the signs can be omitted. For even d, let us define the k-*length *of f as the smallest number $m \in \mathbb{N} \cup \{\infty\}$ such that f has a k-Waring decomposition of length m, with all signs being positive. We denote it by $\operatorname{len}_k f$ and understand it as ∞, whenever there is no such decomposition. In the case $d = 2$, $\operatorname{len} f := \operatorname{len}_2 f$ is called the* Sum-of-Squares length*, or simply the* length *of f. The k-Waring rank of a* generic *kd-form in n variables is denoted $\operatorname{rank}_k^{\circ}(n, kd)$.*

2.3. **Powers of linear forms.** The case $k = 1$ of powers of linear forms is comparatively well-understood. A classical uniqueness result for cubic forms of very low rank is known due to Jennrich (via Harshman [31]). There exist efficient methods to extract the linear forms, c.f. Anandkumar, Ge, Hsu, Kakade and Telgarsky [3].

Theorem 2.3. *(cf. e.g.* [42],[3]*) There exists an algorithm that, on input $n \in \mathbb{N}$ and forms f_2, f_3 of degrees 2 and 3, respectively, computes the solution to the*

following problem: If f_2, f_3 have a POF decomposition

$$f_d = \sum_{i=1}^{m} \lambda_i \ell_i^d \tag{2.11}$$

such that $\ell_1, \dots, \ell_m$ are linearly independent and $\lambda_1, \dots, \lambda_m \in \mathbb{R} \setminus \{0\}$, then compute $(\ell_1, \lambda_1), \dots, (\ell_m, \lambda_m)$. Under these conditions, (2.11) is the unique minimum rank POF decomposition of (f_2, f_3) and the only POF decomposition with linearly independent addends.

Proof. The proof is deferred to Appendix A. □

3. Basic Algorithm for PoF decomposition

3.1. **Overview of ideas and techniques.** The main result of this paper is a recovery algorithm for the addends and weights of a third order powers-of-forms decomposition, as described in Section 2.2. It is simultaneously also an algorithmic proof of uniqueness of the decomposition, and can thus be seen as a generalization of the classical result, Theorem 2.3, commonly attributed to Jennrich [31]. Some of the conditions impose implicit constraints on the rank of the POF decomposition. Section 4 discusses these implications in detail and Section 4.2 proves uniqueness of the POF decomposition for some concrete examples.

The algorithm combines two simple ideas: First, we aim to recover the space $\langle q_1, \dots, q_m \rangle$ spanned by the addends. Note that this is a trivial task for $k = 1$, but for $k \geq 2$ it is not. Then, given a basis $u_1, \dots, u_m$ for the space $\langle q_1, \dots, q_m \rangle$, note that sometimes it is possible to reduce the k-Waring decomposition problem to a 1-Waring decomposition problem. Let us start by explaining this second idea:

Second idea: Reduction to $k = 1$. $\mathbb{R}[q_1, \dots, q_m]$ is an algebra graded by $\frac{1}{k}$ times the degree and the kernel of the graded algebra homomorphism

$$\varphi \colon \mathbb{R}[Y_1, \dots, Y_m] \to \mathbb{R}[q_1, \dots, q_m], Y_1 \mapsto u_1, \dots, Y_m \mapsto u_m \tag{3.1}$$

is the ideal $I_{\mathrm{rel}}(u_1, \dots, u_m)$ of algebraic relations of $u_1, \dots, u_m$, which, via a change of coordinates, translates to the ideal of relations of $q_1, \dots, q_m$. If $I_{\mathrm{rel}}(q_1, \dots, q_m)$ does not contain forms of degree at most 3, then the restriction $\varphi_{\leq 3}$ of φ to $\mathbb{R}[Y_1, \dots, Y_m]_{\leq 3}$ is an invertible linear map onto its image $\mathbb{R}[q_1, \dots, q_m]_{\leq 3}$. The inverse map $\varphi_{\leq 3}^{-1}$ must map the k-forms $q_1, \dots, q_m$ in $X_1, \dots, X_n$ to some linear forms $\ell_1, \dots, \ell_m$ in $Y_1, \dots, Y_m$. One easily sees that for $d \in \{1, 2, 3\}$:

$$g_d := \varphi_{\leq 3}^{-1}(f_d) = \sum_{i=1}^{m} \ell_i^d \tag{3.2}$$

admit a joint decomposition as powers of *linear* forms. From that, a classical algorithm based on eigenvalue decomposition can be used, which is described in Theorem 2.3 and Appendix A. Note that the inverse of $\varphi_{\leq 3}$ can be computed: Since the d-fold products $(u^\alpha)_{|\alpha|=d}$ of entries of u form a basis of $\mathbb{R}[U]_d$ for $d = 1, 2, 3$, there exist unique coefficients $(c_\alpha)_{|\alpha|=d}$ such that

$$f_d = \sum_{|\alpha|=d} c_\alpha u^\alpha, \quad (d = 1, 2, 3), \tag{3.3}$$

which can be obtained by linear system solving. Then, $g_d = \sum_{|\alpha|=d} c_\alpha Y^\alpha$.

It is easy to see that the invertibility condition is a generic property at least as long as $m \leq n + 1$, but e.g. for quadratics $q_1, \dots, q_m \in \mathbb{R}[X]_2$, it also holds for some $m = \Theta(n^2)$. This is discussed in Section 4.1.

First idea: Space recovery. To recover a basis of the space $\langle q_1, \dots, q_m \rangle$, we make heuristic use of the Gram spectrahedron. By Proposition 2.1, there is a superspace U_F of $\langle q_1, \dots, q_m \rangle$ associated with every face F of $\operatorname{Gram}(f_2)$ containing the Gram matrix $G(q)$ of the representation $f_2 = \sum_{i=1}^m q_i^2$. In particular, one of these subspaces is equal to $\langle q_1, \dots, q_m \rangle$. It corresponds to the supporting face of $G(q)$.

It is not clear whether the faces containing $G(q)$ are accessible to us from input f_2, f_3. Fortunately, there are many cases where $\operatorname{Gram}(f_2)$ has a particularly simple structure. The simplest possible case is when f_2 is uniquely Sum-of-Squares representable. Then, $\operatorname{Gram}(f_2) = \{G(q)\}$ is a singleton. It suffices to compute the unique Gram matrix $G = G(q)$ of f_2 and its image will give the space of $q_1, \dots, q_m$. The second simplest case is when $G(q) \in \operatorname{relint} \operatorname{Gram}(f_2)$: Then, while there might be several nonequivalent Sum-of-Squares representations, the space $\langle q_1, \dots, q_m \rangle$ is still accessible to us, since we can compute a relative interior point of $\operatorname{Gram}(f_2)$ with an interior point solver for SDPs. If f_2 is constructed from (not too many) *generic* addends $q_1, \dots, q_m$, then $G(q) \in \operatorname{relint} \operatorname{Gram}(f_2)$ is in fact equivalent to $\operatorname{Gram}(f_2) = \{G(q)\}$, cf. Proposition 4.4. In both these cases, it holds $\operatorname{sosupp} f_2 = \langle q_1, \dots, q_m \rangle$ by Proposition 2.1(f). In all other cases, note that one may still take the potentially larger space $\operatorname{sosupp} f_2$ as an "upper approximation" for the space $\langle q_1, \dots, q_m \rangle$, and hope for the best.

To justify that the approach is reasonable, we will show that there are sufficiently many choices of $q_1, \dots, q_m$, such that their second order power sum f_2 is uniquely Sum-of-Squares representable. This is done in Section 4.

3.2. **Algorithms.** The procedure to recover the POF decomposition is described in Algorithm 1. The following theorem is a uniqueness result for POF decomposition, derived as a consequence. Note that for the first read, it is instructive to have the case in mind where $f_2 = \sum_{i=1}^m \lambda_i q_i^2$ is uniquely Sum-of-Squares representable. In this case, $N = m$ in both Theorem 3.1 and Algorithm 1, and it holds $U = \langle q_1, \dots, q_m \rangle$. The condition $\dim \mathbb{R}[U]_{3k} = \binom{m+2}{3}$ is then equivalent to $I_{\mathrm{rel}}(q_1, \dots, q_m)_3 = \{0\}$, which, according to [6, Section 6.4], is satisfied for generic $q_1, \dots, q_m$ with $m = \Theta(n^2)$. Note that a tentative implementation of Algorithm 1, with an example Julia notebook, can be found on GitHub. See [**?**].

Theorem 3.1. *Let $k \in \mathbb{N}$ and let f_2, f_3 be forms of degree $2k$ and $3k$, respectively. Denote $U := \operatorname{sosupp} f_2$ and $N := \dim U$. Assume that the graded component $\mathbb{R}[U]_{3k}$ has dimension $\binom{N+2}{3}$. Then, f_2 and f_3 have at most one joint POF decomposition*

$$f_d = \sum_{i=1}^m \lambda_i q_i^d, \qquad d = 2, 3, \tag{3.4}$$

with positive weights $\lambda_1, \dots, \lambda_m > 0$, $m \in \mathbb{N}$ and linearly independent $q_1, \dots, q_m$. Furthermore, if such a decomposition exists, then Algorithm 1 computes it efficiently and it is the unique minimum rank POF decomposition of (f_2, f_3).

Proof. Assume there are two distinct POF decompositions, the left one of which had linearly independent addends,

$$\sum_{i=1}^m \lambda_i q_i^d = f_d = \sum_{i=1}^{m'} \mu_i p_i^d, \quad d = 2, 3, \tag{3.5}$$

with positive weights λ_i, μ_i and $m, m' \in \mathbb{N}$. Then by Proposition 2.1, it holds that

$$\langle q_1, \dots, q_m \rangle \subseteq U \supseteq \langle p_1, \dots, p_{m'} \rangle. \tag{3.6}$$

The linearly independent system $q_1, \dots, q_m$ can therefore be extended to a basis $u = (q_1, \dots, q_m, u_{m+1}, \dots, u_N)$ of U. By assumption on the dimension of $\mathbb{R}[U]_3$, there are no algebraic relations of degree 3 between linearly independent elements

of U. Since these relations form a homogeneous ideal, there are also no relations of degree *at most* 3. For the evaluation map

$$\varphi\colon \mathbb{R}[Y_1,\dots,Y_N] \to \mathbb{R}[u_1,\dots,u_N], Y_i \mapsto u_i \quad (i=1,\dots,N), \tag{3.7}$$

which sends forms of degree $d \in \mathbb{N}_0$ to forms of degree kd, the restriction $\varphi_{\leq 3}$ to $\mathbb{R}[Y]_{\leq 3}$ is therefore invertible. The inverse $\varphi_{\leq 3}^{-1}$ maps f_2 and f_3 to quadratic and cubic forms g_2 and g_3, respectively, which admit decompositions

$$g_d = \sum_{i=1}^{m} \lambda_i X_i^d = \sum_{i=1}^{m'} \mu_i \ell_i^d, \qquad (d=2,3) \tag{3.8}$$

where $\ell_i := \varphi_{\leq 3}^{-1}(p_i)$. However, Theorem 2.3 shows uniqueness of the rank-m POF decomposition for g_2 and g_3. Precisely, Theorem 2.3 implies that $m' \geq m$ and if $m = m'$, then, up to reordering, $\lambda_i = \mu_i$ and $X_i = \ell_i$ for all $i \in \{1,\dots,m\}$. Substituting back via φ yields $q_i = p_i$ for all $i \in \{1,\dots,m\}$. □

Algorithm 1 Semidefinite algorithm for powers-of-forms decomposition.

Input: $k \in \mathbb{N}$ and forms $f_2 \in \mathbb{R}[X]_{2k}, f_3 \in \mathbb{R}[X]_{3k}$.

Assumptions:

(1) f_d have a joint POF decomposition $f_d = \sum_{i=1}^m \lambda_i q_i^d$ for $d = 2,3$, where $q_1,\dots,q_m \in \mathbb{R}[X]_k$ are linearly independent k-forms, $m \in \mathbb{N}$ and $\lambda_1,\dots,\lambda_m \in \mathbb{R}_{>0}$.

(2) For $U := \operatorname{sosupp} f_2$ and $N := \dim U$, $\mathbb{R}[U]_3$ has dimension $\binom{N+2}{3}$.

Output: $\{(q_1,\lambda_1),\dots,(q_m,\lambda_m)\}$

Procedure:

1: Use f_2 and an interior point SDP solver to compute some basis $u = (u_1,\dots,u_N)$ of U. This can be done by computing some $G \in \operatorname{relint}\operatorname{Gram}(f)$ and a basis of $\operatorname{im} G$, see Proposition 2.1. Cf. Appendix B for the SDP formulation.

2: The linear system

$$f_d = \sum_{|\alpha|=d} c_\alpha u^\alpha, \qquad c_\alpha \in \mathbb{R}, \qquad (\alpha \in \mathbb{N}_0^N, |\alpha| = d)$$

has a unique solution $c_d = (c_\alpha)_{|\alpha|=d} \in \mathbb{R}[Y_1,\dots,Y_N]_d$ for both $d=2$ and $d=3$. Compute it and set

$$g_d := \sum_{|\alpha|=d} c_\alpha Y^\alpha.$$

Note that $g_d = \varphi_{\leq 3}^{-1}(f_d)$, where

$$\varphi_{\leq 3}\colon \mathbb{R}[Y_1,\dots,Y_N]_{\leq 3} \to \mathbb{R}[u_1,\dots,u_N]_{\leq 3}, Y_i \mapsto u_i \qquad (i=1,\dots,N).$$

3: For degree reasons and since the map $\varphi_{\leq 3}$ is the restriction of an algebra homomorphism, there exist unique linearly independent linear forms $\ell_1 = \varphi_{\leq 3}^{-1}(q_1),\dots,\ell_m = \varphi_{\leq 3}^{-1}(q_m)$ such that $g_d = \sum_{i=1}^m \lambda_i \ell_i^d$ for $d = 2,3$.

4: Compute $\{(\ell_1,\lambda_1),\dots,(\ell_m,\lambda_m)\}$ with the algorithm from Theorem 2.3.

5: **return** $\{(\varphi_{\leq 3}(\ell_1),\lambda_1),\dots,(\varphi_{\leq 3}(\ell_m),\lambda_m)\}$.

4. Interpretation of Requirements

4.1. **Real Geometry Viewpoint.** Let us try to understand the conditions of Theorem 3.1 in terms of geometrical properties of the addends $(q_1,\dots,q_m)\in\mathbb{R}[X]_k^m$. Throughout this section, the addends are assumed to be generic forms. Our goal is twofold: On one hand, we want to formulate geometric criteria that are sufficient for the recovery. On the other hand, we want to justify our seemingly heuristic usage of the Gram spectrahedron. To this end, we will, for instance, examine the values of m, for which uniquely representable sums of squares occur for typical choices of addends. In this section, we will prove Corollary 1.3, Corollary 1.4 and Corollary 1.5. Note that Corollary 1.3 is actually a direct consequence of Theorem 3.1:

Proof of Corollary 1.3. If $f_2=\sum_{i=1}^m\lambda_i q_i^2$ is uniquely Sum-of-Squares representable, then the space $U:=\operatorname{sosupp} f_2$ equals $\langle q_1,\dots,q_m\rangle$ by Proposition 2.1. Since $q_1,\dots,q_m$ do not satisfy any algebraic relations of degree 3, the space $\mathbb{R}[U]_3$ has dimension $\binom{m+2}{3}$ and Theorem 1.1 yields the claim. □

In fact, one sees that Theorem 3.1 yields uniqueness of the POF decomposition $f_d=\sum_{i=1}^m\lambda_i q_i^d, d=2,3$ under these two simplified conditions:

- There are no algebraic relations of $q_1,\dots,q_m$ of degree at most 3. (4.1)
- $\operatorname{sosupp}(\sum_{i=1}^m q_i^2)=\langle q_1,\dots,q_m\rangle$ (4.2)

The positive weights $\lambda_1,\dots,\lambda_m$ matter for neither of these conditions, which is why we omit them throughout the section. Let us now discuss (4.1) and (4.2).

First condition: Algebraic relations of general k-forms. Condition (4.1) is the easier one: General choices of $q_1,\dots,q_m$ will not have any algebraic relations at all as long as $m\le n$. However, since we are only interested in degree-3 relations, the maximum value $\beta_k(n)$ of k-forms $q_1,\dots,q_{\beta_k(n)}\in\mathbb{R}[X]_k$ such that $I_{\mathrm{rel}}(q_1,\dots,q_{\beta_k(n)})_{\le 3}=\{0\}$ could potentially be much higher, with the obvious bound $\beta_k(n)\le\binom{n+k-1}{k}$ from linear relations. In [6, Section 6.4], the authors give a combinatorial proof that $\beta_2(n)=\Theta(n^2)$. I am not aware of any other reference for this statement. A numerical study combined with OEIS suggests that $\beta_2(n)$ is at least $\lceil\frac{(n+2)(n+1)}{6}\rceil$, with the lower bound being obtained from the explicit instance

$$q_{ijk}:=(X_i+X_j+X_k)^2,\qquad (i\le j\le k, i+j+k\equiv_n 0) \tag{4.3}$$

Note that this explicit instance would give a lower bound for generic rank-r quadratics $q_1,\dots,q_m$ of all ranks $r\in\{1,\dots,n\}$, by choosing $\mathcal{D}:=\mathcal{D}_r$ as the class of quadratic forms of rank r in the subsequent Proposition 4.1.

Proposition 4.1. *Let $K\in\{\mathbb{R},\mathbb{C}\}$ and $m,n,k,d\in\mathbb{N}_0$. Let $\mathcal{D}\subseteq K[X]_k$ an irreducible variety containing m distinct forms that do not satisfy any relations of degree kd. Let $q_1,\dots,q_m$ general in $\mathcal{D}$. Then also $I_{\mathrm{rel}}(q_1,\dots,q_m)_{kd}=\{0\}$.*

Proof. Let $d\in\mathbb{N}$ and $N:=\{\alpha\in\mathbb{N}_0^m\mid|\alpha|=d\}$. We show that there are no algebraic relations in degree kd over $K=\mathbb{C}$, which clearly also shows the claim over $K=\mathbb{R}$. Consider the variety

$$W_{m,d}=\{[\lambda:q]\in\mathbb{P}(\mathbb{C}^N\times S^k(\mathbb{C}^n)^m)\mid\sum_{\alpha\in\mathbb{N}_0^m,|\alpha|=d}\lambda_\alpha q^\alpha=0\} \tag{4.4}$$

consisting of pairs of $(q_1,\dots,q_m)$ and the coefficients λ of relations in between them. Let π denote the projection to the q-coordinates and consider any tuple of forms $q=(q_1,\dots,q_m)$. Then the fiber $\pi^{-1}(\{q\})$ is a (projective) subspace of $\mathbb{P}(\mathbb{C}^N)$ corresponding to the algebraic relations of $q_1,\dots,q_m$ in degree d. The set of points

q such that $\pi^{-1}(\{q\})$ is nonempty is Zariski closed. Thus if we find a specific point q such that $\pi^{-1}(\{q\})$ is empty, it will be empty on a Zariski open neighbourhood of q. But such a specific sequence $q = (q_1, \ldots, q_m) \in \mathcal{D}$ exists by assumption. □

Conjecture 4.2. *The family*

$$q_{ijk} := (X_i + X_j + X_k)^2, \qquad (i \leq j \leq k, i + j + k \equiv 0 \mod n) \tag{4.5}$$

of $m = \lceil \frac{(n+2)(n+1)}{6} \rceil$ *quadratics in* n *variables* $X = (X_1, \ldots, X_n)$ *does not satisfy any algebraic relations of degree* 3.

Evidence. Verified on a computer for $n \in \{7, \ldots, 12\}$. □

Remark 4.3. *For* $m = n + 1$, $k \geq 2$, $K \in \{\mathbb{R}, \mathbb{C}\}$ *and general* $q_1, \ldots, q_m \in K[X]_k$, *the ideal of relations* $I_{\mathrm{rel}}(q_1, \ldots, q_m) = (f)$ *is principal and generated by some* $f \in K[Y_1, \ldots, Y_m]$ *of degree* k^n. *Indeed, the polynomial map*

$$\underline{q} \colon K^n \to K^m, x \mapsto (q_1(x), \ldots, q_m(x)) \tag{4.6}$$

has n*-dimensional image. The degree of* f *is the degree of the variety* $V(f) \subseteq \mathbb{C}^m$, *which is determined by intersecting* $V(f)$ *with a general subspace* H *of dimension* 1. *Denote by* $\mathcal{L}$ *the* n*-dimensional space of linear equations defining* H. *Choose a basis* $\mathcal{L} = \langle \ell_1, \ldots, \ell_n \rangle$ *where* $\ell_1, \ldots, \ell_n \in \mathbb{C}[Y]_1$. *Pulling back via* $\underline{q}$ *yields a system of* n *quadratic forms* $\ell_i(q_1, \ldots, q_m) \in \mathbb{C}[X]_2$, $i \in \{1, \ldots, n\}$. *By Bézout's theorem and genericity of* H *and* $q_1, \ldots, q_m$, *we obtain that this system has* k^n *(complex) solutions, so* $\deg(f) = k^n$. *For all* $n \geq 2, k \geq 2$, *this means that* $\beta_k(n) \geq n + 1$.

Second condition: Unique Sum-of-Squares representations. Condition (4.2) is more interesting. The space sosupp f_2 is hard to analyze, unless it equals $\langle q_1, \ldots, q_m \rangle$. For generic $q_1, \ldots, q_m$, their second order power sum $f_2 = q_1^2 + \ldots + q_m^2$ will have length m as long as $m \leq \mathrm{rank}_k^\circ(n, 2k)$, cf. Definition 2.2. Thus, $q_1, \ldots, q_m$ form a **minimum** length Sum-of-Squares representation of f_2. Therefore, the Gram matrix $G(q)$ associated with this representation lies on the boundary of $\mathrm{Gram}(f)$, by Proposition 2.1. On the other hand, also by Proposition 2.1, the Gram matrices that have sosupp f_2 as their image are precisely the relative interior points of $\mathrm{Gram}(f_2)$, which correspond to **maximum** length Sum-of-Squares representations of f_2 (with linearly independent addends). Condition (4.2) is thus saying that the boundary point $G(q)$ is also a relative interior point of $\mathrm{Gram}(f_2)$. This is only possible if $\mathrm{Gram}(f_2)$ is a singleton and therefore if f_2 is uniquely Sum-of-Squares representable. Proposition 4.4 collects this easy fact for future reference.

Proposition 4.4. *Let* $k, m \in \mathbb{N}$ *with* $m \leq \mathrm{rank}_k^\circ(n, 2k)$ *and* $q_1, \ldots, q_m \in \mathbb{R}[X]_k$ *be general* k*-forms. Denote* $f_2 = \sum_{i=1}^m q_i^2$. *Then* f_2 *is uniquely Sum-of-Squares representable, if and only if* $\mathrm{sosupp}\, f_2 = \langle q_1, \ldots, q_m \rangle$.

Now, under which geometrical conditions on the variety of $q_1, \ldots, q_m$ does $\sum_{i=1}^m q_i^2$ have a unique Sum-of-Squares representation? We write $V_\mathbb{R} := V \cap \mathbb{R}^n$ for the set of real points of some affine variety $V \subseteq \mathbb{C}^n$.

Reminder 4.5. *A subvariety* $V \subseteq \mathbb{C}^n$ *has dense real points* $V_\mathbb{R} \subseteq \mathbb{R}^n$ *if and only if every irreducible component of* V *contains a real point that is smooth in* V. *In particular, an irreducible nonsingular subvariety* $V \subseteq \mathbb{C}^n$ *has dense real points* $V_\mathbb{R}$ *if and only if it contains a real point.*

Proposition 4.6 (Corollary of Bertini's theorem, cf. [32, II 8.4(d)])**.** *Fix* $k, m \in \mathbb{N}$. *If* $q_1, \ldots, q_m \in \mathbb{C}[X_1, \ldots, X_n]_k$ *are general and* $m \leq n - 1$, *then* $I := (q_1, \ldots, q_m)$ *is a radical ideal, and its variety* $V(I)$ *in* $\mathbb{P}(\mathbb{C}^n)$ *is of pure codimension* m. *If, in addition,* $m \leq n - 2$, *then* I *is a prime ideal and* $V(I)$ *is smooth.*

Lemma 4.7. *Let* $m, k \in \mathbb{N}$. *The following hold:*

(a) *Let $q_1, \dots, q_m \in \mathbb{R}[X]_k^m$. Assume that $I := (q_1, \dots, q_m)$ is a radical ideal and $V_{\mathbb{R}}(I)$ is dense in $V(I)$. Then, for $f_2 := \sum_{i=1}^m q_i^2$, it holds that* $\operatorname{sosupp} f_2 = \langle q_1, \dots, q_m \rangle$ *and f_2 has dual nondegenerate Sum-of-Squares support.*

(b) *Let $q_1, \dots, q_m \in \mathbb{R}[X]_k^m$ be general forms satisfying the assumption of (a) and let $m \leq \operatorname{rank}_k^\circ(n, 2k)$. Then f_2 is uniquely Sum-of-Squares representable.*

Proof. Claim (a) follows immediately from Corollary B.3, proven in Appendix B. Everything except dual nondegeneracy can also be seen directly: Let $N \in \mathbb{N}_0$ and $p_1, \dots, p_N \in \mathbb{R}[X]_k$ such that $\sum_{i=1}^N p_i^2 = \sum_{i=1}^m q_i^2$. Evaluating this identity in some $x \in V_{\mathbb{R}}(I)$ yields that $p_1, \dots, p_m$ vanish on $V_{\mathbb{R}}(I)$. Since $V(I)$ has dense real points and I is radical, $p_1, \dots, p_m$ must lie in $I_k = \langle q_1, \dots, q_m \rangle$. Thus $\langle q_1, \dots, q_m \rangle$ equals $\operatorname{sosupp} f_2$. Claim (b) follows from (a) using Proposition 4.4. □

Corollary 4.8 (Restatement of Corollaries 1.4 and 1.5)**.** *Let $q_1, \dots, q_m \in \mathbb{R}[X]_k^m$ general satisfying one of these properties:*

(1) $m \leq n-2$ and $V_{\mathbb{R}}(q_1, \dots, q_m)$ contains a nonzero point, or

(2) $m = n-1$ and all lines in the affine cone $V(q_1, \dots, q_m)$ are real.

Then, $\sum_{i=1}^m q_i^2$ is uniquely Sum-of-Squares representable, with dual nondegenerate Sum-of-Squares support. In particular, Algorithm 1 recovers $\{q_1, \dots, q_m\}$ from input $\sum_{i=1}^m q_i^2$ and $\sum_{i=1}^m q_i^3$. For fixed $m \leq n-1$, condition (a) or (b), respectively, are satisfied on a Euclidean open subset of $\mathbb{R}[X]_k^m$.

Proof. **Case (1):** Since $m \leq n-2$, Bertini's Theorem 4.6 guarantees that $I = (q_1, \dots, q_m)$ is a prime ideal and the affine cone $V(I)$ is smooth and irreducible. Thus, by 4.5, the condition of Lemma 4.7 is satisfied. From the Poincaré-Miranda theorem C.1, it easy to see that $V(q_1, \dots, q_m)$ contains a real nonzero point for a Euclidean open subset of all k-forms $q_1, \dots, q_m$. This is elaborated in Appendix C.

Case (2): Since $m = n-1$, Bertini's theorem 4.6 yields that $I = (q_1, \dots, q_m)$ is a radical ideal and $V(I)$ is a union of finitely many lines. Thus the condition of Lemma 4.7 is met if and only if all those lines are real. If the affine cone $V(I)$ has the maximum number of k^m real lines given by the Bézout bound for some specific choice of $q_1, \dots, q_m$, then so it does in a neighbourhood: Indeed, if there was a curve $q(t)$ through $q(0) = (q_1, \dots, q_m)$ such that $V(q(t))$ had nonreal lines for each $t \neq 0$, then a single real line of $V(q(0))$ would have to branch into two complex conjugate lines. This is not possible, since the specific system at $t = 0$ already attains the Bézout bound. For the specific choice, one may choose $q_i \in \mathbb{R}[X_i, X_n]_k$ as X_n-homogenizations of univariate polynomials in X_i with k distinct real zeros. □

4.2. **Numerical experiments with trace-free quadratics.** This section conducts a numerical study on random quadratics. Instances $q = (q_1, \dots, q_m)$ of random quadratic forms are sampled such that $q_1, \dots, q_m$ are iid. Subsequently, we use semidefinite programming to determine the dimension of $\operatorname{sosupp}(\sum_{i=1}^m q_i^2)$ and the kernel of some relative interior point of the dual SDP from Appendix B. The quadratics are chosen from a trace-free distribution (see Definition 4.9), to avoid choosing positive definite forms. I wish to thank Greg Blekhermann and João Gouveia for this suggestion, as it appears that for these distributions, the probability to get uniquely representable sums of squares behaves surprisingly regular. Precisely, we consider the following distributions:

Definition 4.9. *Let $n \in \mathbb{N}$ and $X = (X_1, \dots, X_n)$.*

(1) We call a random quadratic q in X Gaussian trace-free*, if it is sampled as follows: Choose the entries of a matrix $A \in \mathbb{R}^{n \times n}$ independently at random from the standard normal distribution $\mathcal{N}(0,1)$. Set $q := X^T(A - \operatorname{tr}(A) I_n)X$.*

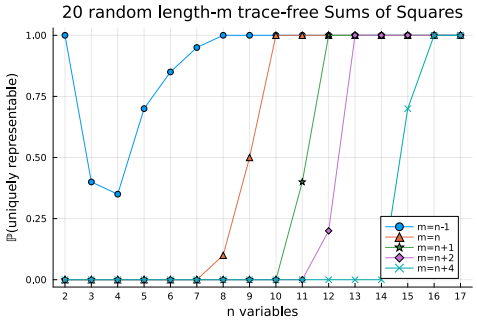

FIGURE 1. The y-axis shows the probability that $\sum_{i=1}^{m} q_i^2$ is uniquely representable, if $q_1, \ldots, q_m$ are iid Gaussian random trace-free quadratics. All probabilities were empirically estimated by sampling the quadratic forms and then solving SDPs. Each data point corresponds to an average over 20 instances. The curves show the behaviour for different relations between m and n.

.

(2) *We call a random quadratic q in X* Gauss-Gramian trace-free*, if it is sampled as follows: Choose the entries of a matrix $A \in \mathbb{R}^{n\times n}$ independently at random from the standard normal distribution $\mathcal{N}(0,1)$. Then, set $q := X^T(A^TA - \operatorname{tr}(A^TA)I_n)X$.*

In addition, some explicit family of $m = \Theta(n^2)$ quadratics is constructed. We verify for the first values $n \in \mathbb{N}$ that the sum of its squares is uniquely representable. We conjecture that this holds true for all $n \in \mathbb{N}$, see Conjecture 4.12. The code and results of all experiments can be found on GitHub, see [67]. Computations were done in `Julia` [10], with the packages `JuMP` [24], `MultivariatePolynomials` [41], `SumOfSquares` [40], [70] and the `Mosek` solver [48]. For the Gaussian trace-free distribution, this is the observed behaviour of the probability $p_{m,n}$ that $\sum_{i=1}^{m} q_i^2$ is uniquely representable and its supporting face in Σ_{2k} is exposed:

(1) If $m = n + r$ for some constant r, then $p_{m,n}$ appears to be an S-shaped curve in n that converges to 1 as $n \to \infty$. This behaviour is depicted in Figure 1. Cf. [67, `data/experiment-3`].
(2) $p_{m,n} > 0$ if $m \leq n - 1$. See the blue curve in Figure 1. This is consistent with the results from Section 4.1. Cf. [67, `data/experiment-3`].
(3) For $m(n) = \lceil \frac{(n+2)(n+1)}{6} \rceil = \Theta(n^2)$, it appears to hold $p_{m(n),2n} \approx 1$, but $p_{m(n),n} \approx 0$. Cf. [67, `data/experiment-2`].

The same qualitative behaviour holds true if the Gaussian trace-free distribution is replaced by the Gauss-Gramian trace-free distribution from Definition 4.9(b). Cf. [67, `data/experiment-2-gramian` and [`data/experiment-3-gramian`]. All of the above statements are empirical observations, based on limited computational experiments in a bounded number of variables. A modest version of the first observation is formulated in the following Conjecture.

Conjecture 4.10. *If $q_1, \ldots, q_n$ are chosen as iid Gaussian random trace-free quadratics, then $\sum_{i=1}^{n} q_i^2$ is uniquely representable with probability $p_n \to 1$ as $n \to \infty$.*

The third observation aligns with Conjecture 4.12, for which we gathered separate numerical evidence. In particular, both suggest that there exist open neighbourhoods of parameters $q = (q_1, \ldots, q_m)$, where m is much larger than n and their sum of squares is uniquely representable. This leads to a natural question: What is the maximum typical length of a uniquely representable sum of squares? Formally:

Question 4.11. *For $n \in \mathbb{N}$, what is the maximum number $m(n)$ of linearly independent quadratics $q_1, \dots, q_{m(n)}$ in n variables such that for all $p_1, \dots, p_{m(n)}$ in some (Euclidean) neighbourhood of $q_1, \dots, q_{m(n)}$, $\sum_{i=1}^{m(n)} p_i^2$ is uniquely representable?*

Conjecture 4.12. *The explicit family of $m = \lceil \frac{(n+2)(n+1)}{6} \rceil$ trace-free quadratics*

$$q_{ijk} := (X_i + X_j + X_k)(Y_i + Y_j + Y_k), \qquad (i \leq j \leq k, i + j + k \equiv_n 0) \tag{4.7}$$

in $2n$ variables $(X, Y) = (X_1, \dots, X_n, Y_1 \dots, Y_n)$ does not satisfy any algebraic relations of degree 3 and

$$\sum_{\substack{i \leq j \leq k \\ i+j+k \equiv_n 0}} q_{ijk}^2 \tag{4.8}$$

is uniquely Sum-of-Squares representable, with nondegenerate dual. The same claim also holds true if q_{ijk} are replaced by $q_{ijk} + \varepsilon p_{ijk}$, where $\varepsilon \in \mathbb{R}$ is sufficiently small and p_{ijk} is some polynomial with support contained in $\{X_i Y_j \mid i, j = 1, \dots, n\}$.

Evidence. In [67, `data/experiment-1`], we checked numerically for $n = 2, \dots, 15$. Note that Conjecture 4.2 would imply that also the specific family here has no algebraic relations up to degree 3, since this one can be reduced to the one from Conjecture 4.2 by substituting $Y_i \mapsto X_i$ for all $i \in \{1, \dots, n\}$. □

5. Applications

Let us now discuss two practical applications of powers-of-forms decomposition.

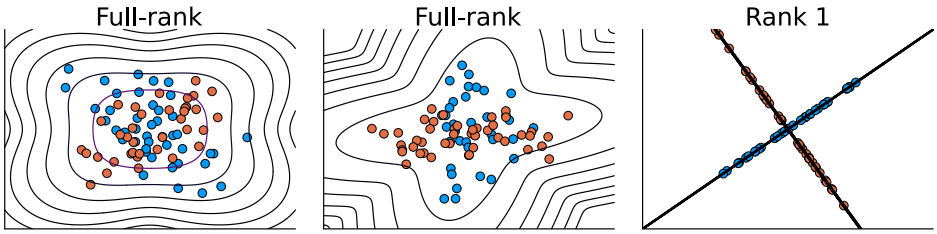

Figure 2. Various centered Gaussian mixture distributions of rank 2. The sample colouring indicates which of the Gaussians was chosen in the sampling process. The pictures on the left and middle show mixtures of full-dimensional Gaussians, as treated in Section 5.1. Here, the contour lines describe the probability density function of the mixture. The rightmost picture shows a mixture on proper subspaces. These are addressed in Section 5.2.

5.1. **Mixtures of centered Gaussians.** The distribution $\mathcal{N}(\mu, \Sigma)$ of a Gaussian random vector on $\mathbb{R}^n$ is parameterized by its mean vector $\mu \in \mathbb{R}^n$ and its symmetric positive definite covariance matrix $\Sigma \in \mathbb{R}^{n \times n}$. A Gaussian random vector is called *centered*, if its mean vector is zero. In that case, all information about its distribution is contained in the quadratic form $q := X^T \Sigma X$. A *Gaussian mixture* Y is a random variable that is sampled as follows: From a box containing m normally distributed random variables $Y_1, \dots, Y_m$, blindly draw one of them (with $\lambda_i \geq 0$ being the probability to draw Y_i, assuming $\sum_{i=1}^m \lambda_i = 1$), and then sample Y_i. We denote $Y = \lambda_1 Y_1 \oplus \dots \oplus \lambda_m Y_m$ for the random variable Y defined by this sampling procedure. Let us now consider a mixture $Y = \lambda_1 \mathcal{N}(0, \Sigma_1) \oplus \dots \oplus \lambda_m \mathcal{N}(0, \Sigma_m)$ of centered Gaussians with covariance forms $q_i = X^T \Sigma_i X$. It turns out that from

sufficiently many samples of Y, it is possible to compute (noisy versions of) the expressions

$$\sum_{i=1}^{m} \lambda_i q_i^d, \qquad d = 0, \dots, D \tag{5.1}$$

where $D \in \mathbb{N}$ is some threshold depending on the order of the number of samples. Up to scalars, the expressions (5.1) are the degree-$2d$ *moment forms* of Y, i.e. the $2d$-homogeneous parts of the *moment generating series* $\mathbb{E}_{\sim Y}[\exp(Y^T X)]$ of Y. The connection between samples, moments and powers-of-forms expressions is explained with lots of details in my doctoral thesis [12, Chapter 3, Introduction].

The goal is to estimate the parameters $q_1, \dots, q_m$ from not too many samples. The *moment problem for mixtures of Gaussians* asks to recover the parameters from (exact) knowledge of the moment forms instead. It can be seen as a coarsening of the statistical estimation problem: To estimate the expression $\sum_{i=1}^m \lambda_i q_i^d$, one needs roughly $\mathcal{O}(\sigma^d)$ iid samples, where the noise from estimation can be made arbitrarily small by taking more samples, σ is a total variance parameter depending on $q_1, \dots, q_m$ and the $\mathcal{O}$-Notation hides e.g. dependency on the dimension n. In order to be efficient with the sample complexity, it is therefore desirable to keep D as small as possible. Previous work of the author with Casarotti, Michałek and Oneto [13] showed that theoretical identifiability of the parameters holds true in our setting, if $D \geq 3$ and m is not too large (roughly $m \in \mathcal{O}(n^{d-1})$). This explains the focus of the present paper on the minimal case $D = 3$. That being said, we are now ready to give a proof of Corollary 1.6:

Proof of Corollary 1.6. Writing $q_i = X^T \Sigma_i X$, the moment forms $\mathcal{M}_{2d}(Y)$ of the Gaussian mixture random variable $Y = \lambda_1 \mathcal{N}(0, \Sigma_1) \oplus \ldots \oplus \lambda_m \mathcal{N}(0, \Sigma_m)$ may be expressed as a convex combination (cf. [12, Chapter 3, Introduction])

$$\mathcal{M}_{2d}(Y) = c_d \sum_{i=1}^{m} \lambda_i q_i^d, \qquad (c_d \in \mathbb{Q}), \tag{5.2}$$

and these are given as input for $d \in \{0, \dots, 3\}$. The combinatorial expression c_d is explicitly known, see [12, Chapter 3, Introduction]. By Corollary 1.4, we know that there exists a Euclidean open subset $\mathcal{U}$ of m-tuples of quadratics where Algorithm 1 recovers the quadratics from their third and second order powers sums. Now, fix some positive definite form p and observe that for $\lambda \in \mathbb{R}_{>0}$ sufficiently large, $\mathcal{U}' := \{q \in \mathcal{U} + \lambda p \mid q_1 \succ 0, \dots, q_m \succ 0\}$ will be a nonempty Euclidean open subset of tuples of positive definite quadratics. On this subset, the following algorithm works:
Choose a new variable Z and compute

$$f_2 = \sum_{i=1}^{m} \lambda_i (q_i - Z)^2, \quad f_3 = \sum_{i=1}^{m} \lambda_i (q_i - Z)^3. \tag{5.3}$$

from the input. This is possible, since the X-homogeneous parts of the power sums from (5.3) correspond to the moments forms $\mathcal{M}_0(Y), \mathcal{M}_2(Y), \dots, \mathcal{M}_6(Y)$. Plug in $Z \mapsto \lambda p$ to obtain an instance where Algorithm 1 succeeds to recover the addends $q_1 - \lambda p, \dots, q_m - \lambda p$, with corresponding weights. Shift back by λp to recover the covariance forms. □

Remark 5.1. *The numerical experiments, which lead to Conjecture 4.10, suggest that the shifting method from the proof of Corollary 1.6 works with asymptotic probability* 1 *in an "average case" framework, where the covariance forms $q_1, \dots, q_m$ are constructed from "random positive definite matrices": Consider $q_1, \dots, q_m$ sampled iid from the distribution $\frac{1}{n^2} X^T A^T A X$, where A is a random $n \times n$ matrix with iid*

Gaussian distributed entries in $\mathcal{N}(0,\sigma^2)$. Here, $\frac{1}{n^2}\operatorname{tr}(A^TA)$, which is the squared Frobenius norm of $\frac{1}{n}A$, will concentrate around the expected value, which is σ^2, with high probability. Thus for large n, $q_1,\dots,q_m$ will all have roughly the same trace. A proxy for this value is algorithmically accessible, since $\sigma^2 \approx \frac{1}{m}\operatorname{tr}(\sum_{i=1}^m q_i)$. A posteriori, this motivates Definition 4.9(b) of Gauss-Gramian quadratic forms, since that will be the distribution of Gaussian random psd forms after shifting by the trace.

5.2. **Learning a union of subspaces.** Learning unions of subspaces is a comparatively new problem that emerged from applications in computer vision and dimensionality reduction techniques in data science ([69],[33],[43],[36]). It assumes that the given data stems from a distribution, whose support is a union of r-dimensional subspaces $U_1,\dots,U_m$. r is known and the objective is to find bases for the subspaces $U_1,\dots,U_m$ from samples of the mixture distribution as input. A union of subspaces is an algebraic variety. E.g., if the subspaces are hyperplanes defined by linear forms, then their union is the zero set of the product of the linear forms. Note that without a distributional assumption, recovering this variety is likely the best one can do, and it can be a hassle to recover the high-degree polynomials describing it from data.

However, assuming that the data on the individual subspaces is Gaussian distributed, it is often possible to recover the subspaces using degree-6 moments of the empirical data. More so, it is then possible to learn the distributions on the subspaces. Indeed, note that this Gaussian subspace learning is a generalized version of the Gaussian mixture problem from Section 5.1, with the difference that Gaussians on subspaces will lead to psd forms $q_1,\dots,q_m$ that are not of full rank.

Corollary 5.2 (Restatement of Corollary 1.7)**.** *For any $n,r\in\mathbb{N}$, $m\le n-1$, there is a Euclidean open subset $\mathcal{U}$ of the problem parameters and an efficient algorithm for the following problem: If $Y_1,\dots,Y_m$ are normally distributed random variables on r-dimensional subspaces $U_1,\dots,U_m$ and $\lambda\in\mathbb{R}^m_{>0}$ with $\sum_{i=1}^m\lambda_i=1$, compute bases for the subspaces $U_1,\dots,U_m$ from the moments of $\lambda_1Y_1\oplus\ldots\oplus\lambda_mY_m$ of degree at most 6.*

Proof. For each $i\in\{1,\dots,m\}$, there exists a unique quadratic form q_i on $\mathbb{R}^n$ such that the restriction of q_i to U_i equals the covariance form of Y_i and the kernel of q_i is the orthogonal complement $U_i^\perp$ of U_i (with respect to the standard inner product). It is then not too hard to see that the even degree moment forms of degree at most 6 of $\lambda_1Y_1\oplus\ldots\oplus\lambda_mY_m$, up to known scalars, attain the form

$$\sum_{i=1}^m \lambda_i q_i^d, \qquad d=0,1,2,3. \tag{5.4}$$

Write $\mathcal{D}_r$ for the class of quadratic forms of rank at most r. If $r=1$, then the q_i are squares of linear forms and we can directly use an algorithm for 1-Waring decomposition, similar to the one from Appendix A. Thus, wlog $r\ge 2$.

Consider first a special instance: For $m\le n-1$, $I=(X_1^2-X_n^2,\dots,X_m^2-X_n^2)$ is a radical ideal: Indeed, if $\mathfrak{P}$ is a minimal prime containing I, then by Krull's Hauptidealsatz, $\mathfrak{P}$ has length at most m. In addition, for each $i\in\{1,\dots,m\}$, we have $X_1-X_n\in\mathfrak{P}$ or $X_1+X_n\in\mathfrak{P}$. Thus, there exists a sign choice $\sigma\in\{\pm1\}^m$ such that $(X_1+\sigma_1X_n,\dots,X_m+\sigma_mX_n)\subseteq\mathfrak{P}$. Since the ideal to the left is prime and of height m, we have equality. We conclude that any minimal prime containing I is of the form $\mathfrak{P}=(X_1+\sigma_1X_n,\dots,X_m+\sigma_mX_n)$. A quick exercise shows that the intersection of those 2^m primes is indeed I. Thus, I is radical. Denote $q_{\text{spec}}:=(X_i^2-X_n^2)_{i=1,\dots,m}$ for those special quadratic forms.

By the Kleiman-Bertini theorem [27, Appendix B, 9.2], the set of rank-r forms $q = (q_1, \dots, q_m)$ such that $V(q_1, \dots, q_m)$ is not of pure codimension m, is closed. Thus, we find a Zariski open neighbourhood $\mathcal{U}$ of q_{spec} such that for all $q \in \mathcal{U}$, all irreducible components of $V(q)$ have codimension m. Consider thus a general subspace $\mathcal{H}$ in $\mathbb{C}^n$ of fitting dimension such that the intersection $\mathcal{H} \cap V(q)$ consists of $\deg V(q)$ many lines (or, projective points). Since $\deg V(q_{\text{spec}})$ attains the Bézout bound, by semicontinuity of the degree, all q in a neighbourhood will satisfy $\deg V(q) = 2^m$, too. For $q = q_{\text{spec}}$, all 2^m projective points in $\mathcal{H} \cap V(q)$ are real. This property must also hold in a Euclidean neighbourhood: Indeed, if a sequence of general $q^{(n)} \in \mathbb{R}[X]_2^m$ existed such that $q^{(n)} \to q_{\text{spec}}$ as $n \to \infty$ and $H \cap V(q^{(n)})$ did contain non-real points, then a pair of complex conjugate non-real solutions (z_n, z_n^*) would degenerate to just one real solution. Since $H \cap V(q)$ has a constant number of 2^m points in a neighbourhood of q_{spec}, this is not possible. As $\mathcal{H}$ was an arbitrary (general) hyperplane, it follows that $V(q)$ has dense real points in a Euclidean neighbourhood $\mathcal{U}'$ of $q = q_{\text{spec}}$.

By Lemma 4.7(b), Algorithm 1 succeeds to recover the addends, if the (weighted) power sums are constructed from elements of $\mathcal{U}'$. Therefore, on the open subset $\mathcal{D}_r \cap (\mathcal{U} + 2X_n^2)$ of quadratics of rank r, the following algorithm works:

(1) Choose a new variable Z and compute
$$\sum_{i=1}^m (q_i - Z)^2 \quad \text{and} \quad \sum_{i=1}^m (q_i - Z)^3. \tag{5.5}$$
from the input.
(2) Plug $Z \mapsto 2X_n$ into the forms from (5.3) to obtain power sums f_2, f_3, whose unique decomposition has addends in $\mathcal{U}$.
(3) Use Algorithm 1 on input f_2, f_3 to compute some $(\hat{q}_1, \lambda_1), \dots, (\hat{q}_m, \lambda_m)$.
(4) Output $\{(\hat{q}_1 + 2X_n^2, \lambda_1), \dots, (\hat{q}_m + 2X_n^2, \lambda_m)\}$.

The set $\mathcal{D}_r \cap \mathcal{U}'$ is open in $\mathcal{D}_r$ and intersects the subset PSD_r of rank-r psd quadratics in the point $q = (X_i^2 - X_n^2)_{i=1,\dots,n-1}$. Every neighbourhood of a point in PSD_r contains points in the interior of PSD_r. This means that $\mathcal{U}' \cap \text{PSD}_r$ contains a nonempty open subset $\mathcal{U}''$ of PSD_r, showing the claim. Note that the weights can be arbitrary positive reals, summing up to 1. □

References

[1] Amendola, C., Faugere, J.-C., and Sturmfels, B. Moment varieties of Gaussian mixtures. *Journal of Algebraic Statistics 7*, 1 (2016).

[2] Améndola, C., Ranestad, K., and Sturmfels, B. Algebraic Identifiability of Gaussian Mixtures. *International Mathematics Research Notices 2018*, 21 (Nov. 2018), 6556–6580.

[3] Anandkumar, A., Ge, R., Hsu, D., Kakade, S. M., and Telgarsky, M. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research* (2012).

[4] Anderson, J., Belkin, M., Goyal, N., Rademacher, L., and Voss, J. R. The more, the merrier: the blessing of dimensionality for learning large gaussian mixtures. In *Proceedings of The 27th Conference on Learning Theory, COLT 2014* (Barcelona, Spain, 2014), pp. 1135–1164,.

[5] Angelini, E., Galuppi, F., Mella, M., and Ottaviani, G. On the number of Waring decompositions for a generic polynomial vector. *Journal of Pure and Applied Algebra 222*, 4 (2018), 950–965.

[6] Bafna, M., Hsieh, T., Kothari, P., and Xu, J. Polynomial-time power-sum decomposition of polynomials. *2022 IEEE 63st Annual Symposium on Foundations of Computer Science (FOCS)* (2022), (to appear).

[7] Bandeira, A. S., Blum-Smith, B., Kileel, J., Perry, A., Weed, J., and Wein, A. S. Estimation under group actions: recovering orbits from invariants, 2017. arXiv:1712.10163 [math.ST].

[8] Barak, B., Kelner, J. A., and Steurer, D. Dictionary learning and tensor decomposition via the sum-of-squares method. In *Proceedings of the Forty-Seventh Annual ACM Symposium*

on Theory of Computing (New York, NY, USA, 2015), STOC '15, Association for Computing Machinery, p. 143–151.

[9] Barvinok, A. *A Course in Convexity*. American Mathematical Society, 2002.

[10] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. Julia: A fresh approach to numerical computing. *SIAM Review 59*, 1 (2017), 65–98.

[11] Blekherman, G., Smith, G. G., and Velasco, M. Sharp degree bounds for sum-of-squares certificates on projective curves. *Journal de Mathématiques Pures et Appliquées 129* (2019), 61–86.

[12] Blomenhofer, A. T. *Gaussian Mixture Separation and Denoising on Parameterized Varieties*. PhD thesis, Universität Konstanz, Konstanz, 2022.

[13] Blomenhofer, A. T., Casarotti, A., Oneto, A., and Michałek, M. Identifiability for mixtures of centered Gaussians and sums of powers of quadratics, 2022. (to appear in Bulletion of LMS). Preprint: arXiv:2204.09356 [math.AG].

[14] Casarotti, A., and Mella, M. From non-defectivity to identifiability. *Journal of the European Mathematical Society* (2022).

[15] Casarotti, A., and Postinghel, E. Waring identifiability for powers of forms via degenerations, 2023.

[16] Chiantini, L., and Ottaviani, G. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications 33*, 3 (2012), 1018–1037.

[17] Chiantini, L., Ottaviani, G., and Vannieuwenhoven, N. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM Journal on Matrix Analysis and Applications 35*, 4 (2014), 1265–1287.

[18] Chiantini, L., Ottaviani, G., and Vannieuwenhoven, N. On generic identifiability of symmetric tensors of subgeneric rank. *Transactions of the American Mathematical Society 369*, 6 (2016), 4021–4042.

[19] Curto, R. E., and di Dio, P. J. Time-Dependent Moments From the Heat Equation and a Transport Equation. *International Mathematics Research Notices* (09 2022). rnac244.

[20] Dasgupta, S. Learning mixtures of Gaussians. In *40th Annual Symposium on Foundations of Computer Science, FOCS '99* (New York, NY, USA, 1999), p. 634–644.

[21] Dasgupta, S., and Schulman, L. A probabilistic analysis of em for mixtures of separated, spherical Gaussians. *Journal of Machine Learning Research 8*, Feb (2007), 203–226.

[22] De Lathauwer, L., Castaing, J., and Cardoso, J.-F. Fourth-order cumulant-based blind identification of underdetermined mixtures. *IEEE Transactions on Signal Processing 55*, 6 (2007), 2965–2973.

[23] di Dio, P. J. The multidimensional truncated moment problem: Gaussian mixture reconstruction from derivatives of moments. *J. Math. Anal. Appl. 517*, 1 (2023), Paper No. 126592, 27.

[24] Dunning, I., Huchette, J., and Lubin, M. JuMP: A modeling language for mathematical optimization. *SIAM Review 59*, 2 (2017), 295–320.

[25] Fonda, A., and Gidoni, P. Generalizing the Poincaré–Miranda theorem: The avoiding cones condition. *Annali di Matematica Pura ed Applicata 195*, 4 (2015), 1347–1371.

[26] Fröberg, R., Ottaviani, G., and Shapiro, B. On the Waring problem for polynomial rings. *Proceedings of the National Academy of Sciences 109*, 15 (2012), 5600–5602.

[27] Fulton, W. *Intersection Theory*. Ergebnisse der Mathematik und ihrer Grenzgebiete. Springer New York, 2012.

[28] Galuppi, F., and Mella, M. Identifiability of homogeneous polynomials and cremona transformations. *Journal für die reine und angewandte Mathematik (Crelles Journal) 2019*, 757 (2019), 279–308.

[29] Garg, A., Kayal, N., and Saha, C. Learning sums of powers of low-degree polynomials in the non-degenerate case. *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)* (2020).

[30] Ge, R., Huang, Q., and Kakade, S. M. Learning mixtures of Gaussians in high dimensions. *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing* (2015).

[31] Harshman, R. Foundations of the parafac procedure: Models and conditions for an "explanatory" multi-modal factor analysis. *UCLA Working Papers in Phonetics 16* (1970).

[32] Hartshorne, R. *Algebraic geometry*, vol. 52. Springer Science & Business Media, 2013.

[33] Hegde, C., Indyk, P., and Schmidt, L. Fast recovery from a union of subspaces. In *Adv. Neural Inf. Proc. Sys. (NIPS* (2016).

[34] Hilbert, D. Letter adressée à m. hermite. 148–153.

[35] Hirschowitz, A., Alexander, J., and Hirschowitz, A. Polynomial interpolation in several variables. *Journal of Algebraic Geometry 4*, 4 (1995), 201–222.

[36] Hong, D., Malinas, R. P., Fessler, J. A., and Balzano, L. Learning dictionary-based unions of subspaces for image denoising. In *2018 26th European Signal Processing Conference (EUSIPCO)* (2018), pp. 1597–1601.

[37] Hsu, D. J., and Kakade, S. M. Learning mixtures of spherical gaussians: moment methods and spectral decompositions. In *Innovations in Theoretical Computer Science, ITCS '13*. Berkeley, CA, USA, 2013, pp. 11–20,.

[38] Kalai, A. T., Moitra, A., and Valiant, G. Efficiently learning mixtures of two Gaussians. In *STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing* (2010), ACM, New York, pp. 553–562.

[39] Landsberg, J. M. *Tensors: Geometry and applications*. American Mathematical Society, 2012.

[40] Legat, B., Coey, C., Deits, R., Huchette, J., and Perry, A. Sum-of-squares optimization in Julia. In *The First Annual JuMP-dev Workshop* (2017).

[41] Legat, B., Timme, S., and Deits, R. Juliaalgebra/multivariatepolynomials.jl: v0.3.18, 07 2021.

[42] Leurgans, S. E., Ross, R. T., and Abel, R. B. A decomposition for three-way arrays. *SIAM Journal on Matrix Analysis and Applications 14*, 4 (1993), 1064–1083.

[43] Lipor, J., and Balzano, L. Leveraging union of subspace structure to improve constrained clustering. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (2017), ICML'17, JMLR.org, p. 2130–2139.

[44] Liu, A., and Moitra, A. Settling the robust learnability of mixtures of Gaussians. *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing* (2021).

[45] Lundqvist, S., Oneto, A., Reznick, B., and Shapiro, B. On generic and maximal k-ranks of binary forms. *Journal of Pure and Applied Algebra 223*, 5 (2019), 2062–2079.

[46] Ma, T., Shi, J., and Steurer, D. Polynomial-time tensor decompositions with sum-of-squares. In *FOCS* (2016), I. Dinur, Ed., IEEE Computer Society, pp. 438–446.

[47] Moitra, A., and Valiant, G. Settling the polynomial learnability of mixtures of Gaussians. *2010 IEEE 51st Annual Symposium on Foundations of Computer Science* (2010).

[48] MOSEK-ApS. *Semidefinite Optimization — Mosek Optimizer API for Python*, 2019.

[49] Nenashev, G. A note on Fröberg's conjecture for forms of equal degrees. *Comptes Rendus Mathematique 355*, 3 (2017), 272–276.

[50] O'Donnell, R. SOS is not obviously automatizable, even approximately. In *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)* (Dagstuhl, Germany, 2017), C. H. Papadimitriou, Ed., vol. 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, pp. 59:1–59:10.

[51] Pearson, K. Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character 195* (1900), pp. 1–47+405.

[52] Permuter, H. H., Francos, J. M., and Jermyn, I. H. Gaussian mixture models of texture and colour for image database retrieval. In *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03, Hong Kong* (2003), pp. 569–572,.

[53] Ramana, M., and Goldman, A. J. Some geometric results in semidefinite programming. *Journal of Global Optimization 7*, 1 (1995), 33–50.

[54] Regev, O., and Vijayaraghavan, A. On learning mixtures of well-separated gaussians. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS* (2017), IEEE, p. 85–96.

[55] Reynolds, D. A., and Rose, R. C. Robust text-independent speaker identification using gaussian mixture speaker models. *IEEE transactions on speech and audio processing 3*, 1 (1995), 72–83.

[56] Reznick, B. Sums of even powers of real linear forms. *Mem. Amer. Math. Soc. 96*, 463 (1992), viii+155.

[57] Reznick, B. Patterns of dependence among powers of polynomials. In *Algorithmic and quantitative real algebraic geometry (Piscataway, NJ, 2001)*, vol. 60 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.* Amer. Math. Soc., Providence, RI, 2003, pp. 101–121.

[58] Reznick, B. Laws of inertia in higher degree binary forms. *Proc. Amer. Math. Soc. 138*, 3 (2010), 815–826.

[59] Reznick, B. On the length of binary forms. In *Quadratic and higher degree forms*, vol. 31 of *Dev. Math.* Springer, New York, 2013, pp. 207–232.

[60] Reznick, B. Some new canonical forms for polynomials. *Pacific J. Math. 266*, 1 (2013), 185–220.

[61] Reznick, B. Linearly dependent powers of binary quadratic forms. *Pacific J. Math. 303*, 2 (2019), 729–755.

[62] Reznick, B. Equal sums of two cubes of quadratic forms. *International Journal of Number Theory 17*, 03 (2021), 761–786.
[63] Reznick, B., and Tokcan, N. Binary forms with three different relative ranks. *Proc. Amer. Math. Soc. 145*, 12 (2017), 5169–5177.
[64] Sanjeev, A., and Kannan, R. Learning mixtures of arbitrary gaussians. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing* (2001), pp. 247–257,.
[65] Scheiderer, C. Extreme points of gram spectrahedra of binary forms. *Discrete & Computational Geometry 67* (06 2022).
[66] Sylvester, J. J. *Collected works*. Cambridge University Press, Cambridge, 1904.
[67] Taveira Blomenhofer, A. Code and files for: Unique powers-of-forms decompositions from simple Gram spectrahedra, May 2023. Available at github.com/a44l/powers-of-forms.
[68] Terracini, A. Sulle V_k che rappresentano più di $\frac{k(k-1)}{2}$ equazioni di laplace linearmente indipendenti. *Rend. Circ. Mat. Palermo 33* (1912), 176–186.
[69] Wang, Y., Wipf, D., Ling, Q., Chen, W., and Wassell, I. Multi-task learning for subspace segmentation. In *Proceedings of the 32nd International Conference on Machine Learning* (Lille, France, 07–09 Jul 2015), F. Bach and D. Blei, Eds., vol. 37 of *Proceedings of Machine Learning Research*, PMLR, pp. 1209–1217.
[70] Weisser, T., Legat, B., Coey, C., Kapelevich, L., and Vielma, J. P. Polynomial and moment optimization in julia and jump. In *JuliaCon* (2019).

Appendix A. Powers of Linear forms

Theorem A.1 (Restatement of Theorem 2.3)**.** *There exists an efficient algorithm that, on input $m, n \in \mathbb{N}$ and forms f_2, f_3 of degrees 2 and 3, respectively, computes the solution to the following problem: If f_2, f_3 have a POF decomposition*

$$f_d = \sum_{i=1}^{m} \lambda_i \ell_i^d \tag{A.1}$$

such that $\ell_1, \dots, \ell_m$ are linearly independent and $\lambda_1, \dots, \lambda_m \in \mathbb{R} \setminus \{0\}$, then compute $(\ell_1, \lambda_1), \dots, (\ell_m, \lambda_m)$. Under these conditions, (A.1) *is the unique minimum rank POF decomposition of (f_2, f_3) and the only POF decomposition with linearly independent addends.*

Algorithmic Proof. First, note that a partial derivative of f_2 in direction $v \in \mathbb{R}^n$ has the form $\partial_v f_2 = 2\sum_{i=1}^m \lambda_i \ell_i(v)\ell_i$. The set $\{\partial_v f_2 \mid v \in \mathbb{R}^n\}$ equals $\langle \ell_1, \dots, \ell_m \rangle$, by linear independence. Thus, we may compute some basis $u = (u_1, \dots, u_m)$ of $U := \langle \ell_1, \dots, \ell_m \rangle$. Then, there exist vectors $a_1, \dots, a_m \in \mathbb{R}^m$ such that $\ell_i = a_i^T u$. Compute now the partial derivative

$$f_v := \frac{1}{3}\partial_v f_3 = \sum_{i=1}^{m} \lambda_i (a_i^T u(v))(a_i^T u)^2 = u^T M_v u \tag{A.2}$$

of f_3 in some random direction $v \in \mathbb{R}^n$. Here, we write $M_v := \sum_{i=1}^m \lambda_i (a_i^T u(v)) a_i a_i^T$. Similarly, the quadratic form f_2 can be written as $f_2 = u^T M u$, where the matrix $M = \sum_{i=1}^m \lambda_i a_i a_i^T \in \mathbb{R}^{m \times m}$ is symmetric and psd. Note that the matrices M and M_v can easily be computed from u, f_2 and f_v. The claim is now that the generalized eigenvalue problem

$$\det(M_v - \mu M) = 0, \qquad \mu \in \mathbb{R} \tag{A.3}$$

has m onedimensional eigenspaces, with corresponding eigenvalues $\mu_i := a_i^T u(v)$. Indeed, since v was chosen at random from some continuous distribution, M_v is of full rank m, with probability 1. The rank of

$$M_v - \mu M = \sum_{i=1}^{m} \lambda_i ((a_i^T u(v)) - \mu) a_i a_i^T \tag{A.4}$$

drops to $m-1$ precisely if $\mu = \mu_j := a_j^T u(v)$ for some $j \in \{1,\dots,m\}$. Hence, these are the eigenvalues. By randomness of v, the eigenvalues are pairwise distinct. Choose eigenvectors $x_1,\dots,x_m$, satisfying the generalized eigenvalue equation

$$M_v x_j = \mu_j M x_j \tag{A.5}$$

Writing this equation out and comparing coefficients with respect to the basis $a_1,\dots,a_m$, we get that $(a_i^T u(v))(a_i^T x_j) = \mu_j (a_i^T x_j)$ and therefore $a_i^T x_j = a_j^T x_j \cdot \delta_{ij}$ for each $i,j \in \{1,\dots,m\}$. Therefore, with $b_j := Mx_j = \lambda_j (a_j^T x_j) a_j$, we recovered a multiple of a_j. It remains to recover the missing multiples and the weights. To this end, note that the values $\mu_j = a_j^T u(v)$ and $b_j^T u(v)$ are known to us, so we can compute $a_j = \frac{\mu_j}{b_j^T u(v)} b_j$ and thus $\ell_i = a_i^T u$. For the weights, solve the linear system

$$f_2 = \sum_{i=1}^m \nu_i \ell_i^2, \qquad \nu_1,\dots,\nu_m \in \mathbb{R}. \tag{A.6}$$

Since the ℓ_i^2 are linearly independent, this system will have the unique solution $\nu_j = \lambda_j$. This concludes the algorithmic part of the proof.

Regarding the uniqueness statement: For any other decomposition $f_d = \sum_{i=1}^{m'} \rho_i l_i^d$ with $m' \in \mathbb{N}$, $d \in \{2,3\}$, linear forms $l_i \in \mathbb{R}[X]$ and $\rho_i \in \mathbb{R}\setminus\{0\}$, one easily sees that the space $U = \{\partial_v f_2 \mid v \in \mathbb{R}^n\}$ must be contained in $\langle l_1,\dots,l_{m'}\rangle$. Since $\dim U = m$, this means that $m' \geq m$, with equality if and only if $l_1,\dots,l_m$ are linearly independent. If $m = m'$, then the two decompositions are therefore both equal to the output of the algorithm. This means they must be equal. □

Appendix B. Sum-of-Squares Pointedness and strict complementarity

Sum-of-Squares representations of some form $f \in \mathbb{R}[X]_{2k}$ may be found algorithmically via the following primal-dual pair of semidefinite programs:

$$\begin{aligned} \text{(Gram)} \qquad & \text{find} && G \succeq 0 \\ & \text{s.t.} && [X]_k^T G [X]_k = f \end{aligned} \tag{B.1}$$

$$\begin{aligned} \text{(Gram)}^* \qquad & \text{minimize} && E(f) \\ & \text{s.t.} && E \in \mathbb{R}[X]_{2k}^\vee \\ & && M_E \succeq 0 \end{aligned} \tag{B.2}$$

Here, for a functional $E \in \mathbb{R}[X]_{2k}^\vee$ we denote by $M_E := (E(X^{\alpha+\beta}))_{|\alpha|=k=|\beta|}$ the so-called *moment matrix* of E. The moment matrix encodes a psd bilinear form $(p,q) \mapsto E(pq)$ associated with E. It is well-known that $E \in \Sigma_{2k}^*$ if and only if M_E is psd. Note that the dual problem has 0 as its optimal value. The set of optimal solutions defines a face

$$C_f = \{E \in \Sigma_{2k}^* \mid E(f) = 0\} \tag{B.3}$$

of the dual cone Σ_{2k}^*. By complementarity, for each optimal pair (G,E) it holds that $M_E \cdot G = 0$. In other words, $\operatorname{im} G \subseteq \ker M_E$. Taking $G \in \operatorname{relint} \operatorname{Gram}(f)$, we see that each $E \in C_f$ satisfies $\operatorname{sosupp} f \subseteq \ker M_E$ by Proposition 2.1. The latter can also be seen directly: If $\lambda > 0$ such that $f - \lambda p^2 \in \Sigma_{2k}$, then $E(p^2) = 0$ for all $E \in C_f$ and thus by the Cauchy-Schwarz inequality applied to the psd bilinear form M_E, $E(ph)^2 \leq E(p^2)E(h^2) = 0$ for each $h \in \mathbb{R}[X]_k$, implying $p \in \ker M_E$. It is easy to see that the space $\ker M_E$ is constant among all $E \in \operatorname{relint} C_f$.

Strict complementarity is the property $\operatorname{im} G = \ker M_E$. If it holds for some pair (G,E) of primal-dual optimal solutions, then it has to hold for all $G \in \operatorname{relint} \operatorname{Gram}(f)$ and $E \in \operatorname{relint} C_f$. In that case, we have $\ker M_E = \operatorname{sosupp} f$ for

all $E \in \operatorname{relint} C_f$ and we say that f has *dual nondegenerate Sum-of-Squares support*. Dual nondegeneracy can be useful from a practical perspective, since (B.2) has nice properties, e.g., a full-dimensional feasible region. Geometrically, it means that the supporting face of f in Σ_{2k} is exposed (by any $E \in \operatorname{relint} C_f$).

Relation to pointed Sum-of-Squares cones. In this section, we will see that in the "geometric" settings of Corollary 1.4 and Corollary 1.5, the second order power sum has dual nondegenerate Sum-of-Squares support. The argument relies on work of Blekherman, Smith and Velasco [11] and examines pointedness of Sum-of-Squares cones in quotient algebras of the polynomial ring.

Proposition B.1. *(cf. Prop. 2.5. in* [11]*). Let $k, n \in \mathbb{N}$ and $I \subseteq \mathbb{R}[X]$ a homogeneous ideal with graded coordinate ring $R = \mathbb{R}[X]/I$ such that*

$$\forall p \in R_k \colon p^2 \in I_{2k} \implies p \in I_k \tag{B.4}$$

Then the following are equivalent:

(a) The cone $\Sigma_{R,2k}$ is pointed*, i.e. it is closed and contains no lines.*
(b) No nontrivial sum of squares of forms of degree k equals zero in R_{2k}.

Proof. **(a) $\implies$ (b):** By contraposition. Let there be $N \in \mathbb{N}$ and $p_1, \ldots, p_N \in R_k$ such that $\sum_{i=1}^N p_i^2 = 0$. Since no nontrivial squares lie in I_{2k} by assumption, it holds that $N \geq 2$. Thus $p_1^2 = -\sum_{i=2}^N p_i^2$ lies both in $\Sigma_{R,2k}$ and $-\Sigma_{R,2k}$. It follows that $\Sigma_{R,2k}$ contains a line.
(b) $\implies$ (a): By assumption, Σ_{2k} cannot contain a line, since otherwise there would be some nonzero $f \in \Sigma_{2k} \cap -\Sigma_{2k}$ and the nontrivial Sum-of-Squares $f + (-f)$ would be zero. It remains to show that Σ_{2k} is closed. Fix a norm on the real vector space R_k and denote

$$K := \{p^2 \in R_{2k} \mid \|p\| = 1\} \tag{B.5}$$

Then K is a compact basis of the cone Σ_{2k}: Indeed, K is compact since it is the image of a compact set under a continuous function. It holds $0 \notin \operatorname{conv} K$ by assumption, since no nontrivial sum of squares is zero in R_{2k}. Therefore the cone generated by K, which is Σ_{2k}, is closed. □

Proposition B.2. *Let $f \in \Sigma_{2k}$. Write $I := (\operatorname{sosupp} f)$ for the homogeneous ideal generated by the Sum-of-Squares support and $R := \mathbb{R}[X]/I$ for the quotient algebra graded by the degree. Consider the statements:*

(i) f has dual nondegenerate Sum-of-Squares support.
(ii) C_f spans $I_{2k}^\perp$.
(iii) The cone $\Sigma_{R,2k}^* \subseteq R_{2k}^\vee$ is full dimensional.
(iv) The cone $\Sigma_{R,2k} \subseteq R_{2k}$ is pointed, i.e. it is closed and contains no lines.
(v) $\forall p_1, \ldots, p_N \in \mathbb{R}[X]_k \colon p_1^2 + \ldots + p_N^2 \in I_{2k} \implies p_1, \ldots, p_N \in I_k$.

Then it holds (v) $\implies$ (iv) $\implies$ (iii) $\iff$ (ii) $\iff$ (i).

Proof. "(v) $\implies$ (iv)": Note that in Proposition B.1, (b) $\implies$ (a) also holds without the assumption (B.4).
"(iv) $\implies$ (iii)": First, assume $\Sigma_{R,2k}$ is pointed in R_{2k}. If $\Sigma_{R,2k}^* \subseteq H$ was contained in a hyperplane H, then by standard properties of the dual, $\Sigma_{R,2k}^{**}$ would contain the line $H^\perp$. Since $\Sigma_{R,2k}$ is closed, this would yield the contradiction $H^\perp \subseteq \Sigma_{R,2k}^{**} = \overline{\Sigma_{R,2k}} = \Sigma_{R,2k}$.
"(iii) $\iff$ (ii)": The quotient map $\pi \colon \mathbb{R}[X]_{2k} \to R_{2k}$ yields a pullback $\pi^* \colon R_{2k}^\vee \to \mathbb{R}[X]_{2k}^\vee, L \mapsto L \circ \pi$. The map π^* is a bijection onto its image

$$I_{2k}^\perp = \{L \in \mathbb{R}[X]_{2k}^\vee \mid L \text{ vanishes on } I_{2k}\}.$$

The image of $\Sigma_{R,2k}^*$ under π^* equals $\Sigma_{2k}^* \cap I_{2k}^\perp$, which is precisely C_f.

"(ii) $\implies$ (i)": Assume C_f spans $I_{2k}^\perp$. One easily sees that $I_{2k}^\perp$ consists precisely of those functionals L such that I is contained in the kernel of the moment matrix $M_L\colon (p,q) \mapsto L(p,q)$ of L. Equality $I = \ker M_L$ holds for general elements L of $I_{2k}^\perp$ and thus also for relative interior points of C_f. Thus f has dual nondegenerate Sum-of-Squares support.
"(i) $\implies$ (ii)": Assume f has dual nondegenerate Sum-of-Squares support. Then any $E \in \operatorname{relint} C_f$ satisfies $\ker M_E = \operatorname{sosupp} f = I_k$. Now, let $L \in I_{2k}^\perp$. Since $\ker M_L \supseteq \operatorname{sosupp} f$, it is easy to see that for any psd matrix B with $\ker B = \operatorname{sosupp} f$, it holds that $M_L + \lambda B$ is psd for all sufficiently large $\lambda \in \mathbb{R}_{>0}$. Since f has dual nondegenerate Sum-of-Squares support, we may choose some $E \in \operatorname{relint} C_f$ and take $B = M_E$ and $\lambda \in \mathbb{R}_{>0}$ such that $M_L + \lambda M_E \succeq 0$. But then $L + \lambda E \in C_f$, since clearly $(L + \lambda E)(f) = 0$ and the moment matrix of $L + \lambda E$ is psd. Thus $L = (L + \lambda E) - \lambda E$ lies in the span of C_f. □

Corollary B.3. *Let $f \in \Sigma_{2k}$ and let I an ideal with $f \in I \subseteq (\operatorname{sosupp} f)$. If I is real radical, then $I = (\operatorname{sosupp} f)$ and f has dual nondegenerate Sum-of-Squares support.*

Proof. Since I is real radical, for all $p_1, \ldots, p_N \in \mathbb{R}[X]_k$ such that $p_1^2 + \ldots + p_N^2 \in I_{2k}$, it holds $p_1, \ldots, p_m \in I$. This shows that in fact $I = (\operatorname{sosupp} f)$. In addition, by Proposition B.2, it also shows that f has dual nondegenerate Sum-of-Squares support. □

Appendix C. Typical regions with singleton Gram spectrahedra

The condition of Corollary 4.8(a) is satisfied on a Euclidean open subset. This can e.g. be seen from the Poincaré-Miranda theorem:

Theorem C.1 (Poincaré-Miranda, cf. [25, Introduction])**.** *Write $\mathcal{H}$ for the parallelepiped spanned by linearly independent vectors $v_1, \ldots, v_n \in \mathbb{R}^n$ and let $f\colon \mathcal{H} \to \mathbb{R}^n, x \mapsto (f_1(x), \ldots, f_n(x))$ a continuous function. Denote*

$$\mathcal{H}_i^1 := \{\sum_{j=1}^m \lambda_j v_j \mid \lambda_j \in [0,1], \lambda_i = 1\}, \quad \mathcal{H}_i^0 := \{\sum_{j=1}^m \lambda_j v_j \mid \lambda_j \in [0,1], \lambda_i = 0\}$$

for $i \in \{1, \ldots, n\}$. Note these are the facets of $\mathcal{H}$. Assume that for each $i \in \{1, \ldots, n\}$, $f_i \leq 0$ on $\mathcal{H}_i^0$, but $f_i \geq 0$ on $\mathcal{H}_i^1$. Then f has a zero on $\mathcal{H}$.

We may now fill the gap that was left in the proof of Corollary 4.8(a):

Proof of the addendum in Corollary 4.8(a). By Theorem C.1, it clearly suffices to find, say, a rectangle $\mathcal{H}$ and quadratics $q_1, \ldots, q_m, q_{m+1}, \ldots, q_n$ such that for each $i \leq m$, $q_i < 0$ on $\mathcal{H}_i^0$ but $q_i > 0$ on $\mathcal{H}_i^1$, as then the condition of Theorem C.1 will be satisfied in a neighbourhood $\mathcal{U}_1 \times \ldots \times \mathcal{U}_n \times \{q_{m+1}\} \times \ldots \times \{q_n\}$ of $(q_1, \ldots, q_m)$. By introducing some new variable Y, one can take e.g. $m = n$, $q_{n+1} := 0$ and $q_i = (X_i + Y)^2 - 2Y^2 \in \mathbb{R}[X_1, \ldots, X_n, Y]$, for $i \in \{1, \ldots, n\}$ where Y is another unknown, and consider the rectangle $\mathcal{H} = [0,1]^n \times \{1\}$ in the affine plane where "$Y = 1$". This corresponds to choosing the basis $e_1 + e_{n+1}, \ldots, e_n + e_{n+1}, e_{n+1}$ in Theorem C.1. Note $\mathcal{H}_i^a = [0,1]^{i-1} \times \{a\} \times [0,1]^{n-i} \times \{1\}$ for $a \in \{0,1\}$. We have $q_i = -1 < 0$ on $\mathcal{H}_i^0$ and $q_i = 2 > 0$ on $\mathcal{H}_i^1$. Thus, the condition $V_{\mathbb{R}}(q_1, \ldots, q_m) \neq \emptyset$ is a typical property. □

CWI, Networks & Optimization, Amsterdam, Science Park 123, NL-1098 XG.
Email address: atb@cwi.nl
URL: cwi.nl/en/people/filipe-alexander-taveira-blomenhofer