SIAM J. APPL. ALGEBRA GEOMETRY
Vol. 9, No. 1, pp. 211–234

On Uniqueness of Power Sum Decomposition*

Alexander Taveira Blomenhofer†

Abstract. We propose an algorithm to compute power sum decompositions, which are motivated by applications in algebraic statistics. Power sum decomposition entails writing forms of degree $d \cdot k$ as a sum of dth powers of k-forms. We show that under certain assumptions, the power sum problem for k-forms can be reduced to the classical case of power sums of linear forms. Semidefinite programming is used to perform this reduction. The semidefinite programming approach allows us to improve the currently best known rank bounds for the problem from $m = \mathcal{O}(n/\log(n))$ to $m = n-1$, in a typical case. An implementation of the algorithm is provided. We complement the theoretical analysis with numerical experiments.

Key words. waring decomposition, Gaussian mixtures, SDP, subspace learning

MSC codes. 14Q30, 15A69, 62H12, 68Q25

DOI. 10.1137/23M1573355

1. Introduction.

Among Ramanujan's many playful activities, there was a famous problem he posed to the *Journal of the Indian Mathematical Society*, which concerned decompositions of a sextic form f as a sum $f = q_1^3 + q_2^3$ of two cubes of quadratics. See [51], [52, p. 326], [61, p.1 and references therein]. Ramanujan was interested in his particular sextic because it admitted many decompositions, and so he asked the reader to find them. In [61], Reznick gives a complete description of the solutions to "$q_1^3 + q_2^3 = q_3^3 + q_4^3$". In this work, we will study the opposite side: power-sum problems, which have a *unique* solution. For instance, can you find quadratic forms q_1 and q_2 in two variables X, Y such that

$$q_1^2 + q_2^2 = 2X^4 - 2X^3Y + 3X^2Y^2 - 8XY^3 + 5Y^4 \quad \text{and}$$
$$q_1^3 + q_2^3 = -9X^5Y + 18X^4Y^2 - 9X^3Y^3 + 9X^2Y^4 - 18XY^5 + 9Y^6? \tag{1.1}$$

The system (1.1) is a *power sum decomposition problem.* These problems will be very difficult when the number of variables and the number of q_i's is large, but already the tiny example from (1.1) requires significant effort to solve.

Power sum decompositions have a wide range of applications, especially sums of powers of linear forms: These play an important role in polynomial optimization [38, section 5], phylogenetics [37], cryogenic electron microscopy [9], and many more. For the case of linear forms, there is a rich body of literature, both regarding theory, e.g., [41], [4], and algorithms, e.g., [1], [16].

*Received by the editors May 17, 2023; accepted for publication (in revised form) October 23, 2024; published electronically March 10, 2025.
https://doi.org/10.1137/23M1573355
Funding: This work was supported by Dutch Scientific Council grant OCENW.GROOT.2019.015 (OPTIMAL).
†Centrum Wiskunde & Informatica, Amsterdam, The Netherlands, and Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark (atb@math.ku.dk).

If the q_i's have degree greater than 1, then we speak of a *higher-order* power sum problem. The higher-order variant is much less understood, but it has recently received increased interest from various communities, for instance from the viewpoints of arithmetic circuit complexity [25] and high-dimensional estimation [26], [7]. The case of cubes of quadratics is particularly important for a specific class of Gaussian mixture problems, which we address in section 6. Cubes are also the smallest power, at which any nontrivial uniqueness result becomes possible; see [14].

Ideally, we would like to have an algorithm, which, given the power sums $f_2 := q_1^2 + \cdots + q_m^2$ and $f_3 := q_1^3 + \cdots + q_m^3$ as input, can produce a solution $q_1, \ldots, q_m$. Formally, a power sum problem can be written as

$$(1.2) \qquad (\mathrm{PoF})_{\underline{f},m,k,\underline{d}} \qquad \begin{array}{ll} \text{given} & f_1, \ldots, f_r \text{ of degrees } k \cdot d_1, k \cdot d_2, \ldots, k \cdot d_r, \\ \text{find} & q_1, \ldots, q_m \in \mathbb{R}[X_1, \ldots, X_n]_k, \\ & \lambda_1, \ldots, \lambda_m \in \mathbb{R}_{\geq 0}, \\ \text{s. t.} & f_d = \displaystyle\sum_{i=1}^m \lambda_i q_i^d, \qquad d = d_1, \ldots, d_r. \end{array}$$

The number m is called the *rank* of a decomposition. Note that the *weights* λ_i do not occur in (1.1), but they are needed in some of the applications. If one wishes to recover both the weights λ_i and the *addends* q_i, then it is necessary to have power sums of at least two different degrees. If $r = 1$ and all weights are set equal to 1, then the problem is called *k-Waring decomposition.* The system (1.1) has a unique solution, given by

$$(1.3) \qquad q_1 = -X^2 - XY + 2Y^2, \quad q_2 = X^2 - 2XY + Y^2.$$

Uniqueness can be seen as a consequence of our main result, Theorem 3.2, and the solution may be computed using our Algorithm 1.

1.1. Organization and overview of contributions. Our main result is an algorithmic uniqueness theorem for sums of cubic and quadratic powers; see Theorem 3.2 and Corollary 3.3. The corresponding algorithm is described in Algorithm 1. We develop the main result in section 3. A `Julia` implementation of Algorithm 1 can be found in the accompanying supplementary file (powers-of-forms.zip [local/web 333KB]) as well the author's git repository; see [13].

Our result may be used for two estimation problems, which are common in machine learning: the parameter estimation problem for mixtures of centered Gaussians; see Corollary 1.1, as well as subspace recovery; see Corollary 1.2. In section 6, we give a detailed description of these problems.

In section 4, we analyze the algorithm in terms of geometric conditions for uniqueness; see Corollary 4.1. In section 5, we conduct a numerical study, suggesting that the algorithm has good performance in an average-case framework. We also examine the range of ranks for which our method can certify uniqueness of the decomposition.

1.2. Related work and techniques. Waring decompositions are a classical problem, dating back to Sylvester [66], Hilbert [31], and Terracini [67]. There is a significant body of literature and results on the case of 1-Waring decomposition, both on the side of uniqueness theorems, e.g., [16], [24], and on the side of algorithms, e.g., [41], [4]. Forms of very small rank have a unique 1-Waring decomposition of minimal rank, at least generically. In fact, there is a well-known algorithm to compute their 1-Waring decompositions if the rank is small and certain nondegeneracy conditions are met. This algorithm is a classical result, but the attribution to a single origin is complicated, since it has been independently rediscovered multiple times. See Theorem 2.3 for a formulation of the algorithm and subsection 2.2 for a discussion of the origins.

Simultaneous Waring decompositions for vectors of forms of (possibly) different degrees have also been studied before, e.g., [68], [8], [6], and [27].

Higher-order Waring decompositions are a classical topic, too, but they are far less understood. Ramanujan's famous question was about an "unexpected" pattern of dependency between powers of forms, i.e., a power-sum with nonunique minimum-rank decomposition. In this spirit, the pioneering work of Reznick gave various results on the potential solutions of certain power-sum decomposition problems; see [55], [56], [62], [57], [58], [59], [60], and [61]. On the opposite end, it was shown by the author in joint work with Casarotti, Michałek, and Oneto [14] that generic uniqueness of power-sum decomposition holds for most ranks, excluding those ranks where the problem is trivially overparameterized. This motivates the search for algorithms, which generalize those known for linear forms. Unfortunately, the case $k \geq 2$ is more complicated for several reasons. The basic algorithm from Theorem 2.3 relies implicitly on the underlying principle of *apolarity.* A point $a \in \mathbb{R}^n$ can be dually interpreted both as a linear form $a^T X \in \mathbb{R}[X]_1$ and also as a directional differentiation operator $\partial_a \colon \mathbb{R}[X]_d \to \mathbb{R}[X]_{d-1}$. These dualities allow us to associate to a d-form f an ideal of equations, which has exactly the points a as solutions such that $a^T X$ contributed to a Waring decomposition of f.

For higher-order Waring decomposition, it is no longer straightforward to interpret the solutions q_i of $f = q_1^d + \cdots + q_m^d$ as points. Currently, the best-known strategies attempt to reduce to the case $k = 1$ via various tricks. A central step in these reductions is to find the space $\langle q_1, \ldots, q_m \rangle$.

One such trick is the method of affine projections of partials (APP), which was used by Garg, Kayal, and Saha [25] and Bafna et al. [7] to obtain polynomial-time recovery procedures for some variants of powers-of-forms decomposition. Their work was strongly motivated and tailored towards the parameter estimation problem for mixtures of centered Gaussians. Earlier work of Ge Huang, and Kakade [26] already gave a recovery algorithm for the parameters of Gaussian mixtures. Among other things, this implied a recovery algorithm for 2-Waring decompositions of sextic forms if $m = \mathcal{O}(\sqrt{n})$ and if the addends are "sufficiently random". Bafna et al. [7] gave an improved algorithm, also for higher-order Waring decompositions. In the case $k = 2$, it allows one to decompose forms of rank up to $m = \mathcal{O}(n/\log(n))$, again assuming sufficient randomness in the solution. To the best of our knowledge, $m = \mathcal{O}(n/\log(n))$ was the best rank bound known to date. We improve it to $m = n - 1$ for recovery from third- and second-order power sums, via our Corollary 4.1. Compared to the previous results, this therefore gives an improvement of the asymptotic order and the constant factors, with an arguably simpler proof. Note that we need slightly more than genericity conditions: Our assumptions are only satisfied on a typical set. This set appears to be quite large, though, as suggested by the numerical experiments in section 5.

1.3. Statistical learning applications.

Learning parameters of centered Gaussian mixtures. The parameter estimation problem for Gaussian mixtures has a rich history, dating back to Pearson [48]. It has now been studied over more than a century in all kinds of flavors, e.g., from the perspective of computer science [18], [63], [19], [45], [35], [33], [5], [53], [43], algebraic geometry [2], [3], moment problems [17], [20], and applications [54], [49]. A mixture of m *centered* Gaussians has a degree-$2d$ moment form (cf. subsection 6.1) proportional to

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

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 ith 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. This allows one to derive an algorithm for parameter estimation for Gaussian mixtures as a corollary of our main result, Theorem 3.2. The details are elaborated in section 6. Let us highlight the result in the following.

Corollary 1.1. *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.

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 [7], [26] is based on a probabilistic analysis and thus implicitly assumes full-rank quadratics.

Corollary 1.2. *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*[1] *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] The parameters are the subspaces together with the means and covariances of the Gaussians.

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 power sum 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 $x \in K^n$. 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 kth graded component of R and $R_{\leq k} := R_0 \oplus \ldots \oplus R_k$ denotes the part of grade at most k.

For a convex cone $C \subseteq U$ in an $\mathbb{R}$-vector space U, the *dual cone* of C is

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

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

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

For subspaces, C^* is thus a subspace of $U^\vee$, which is called the *conormal space* of C. It is commonly denoted $C^\perp$ rather than C^*.

2.1. Sums of squares and 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

$$f = \sum_{i=1}^{N} q_i^2. \tag{2.5}$$

The right-hand side of (2.5) 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 one to write sum-of-squares representations such as (2.5) via *Gram matrices*

$$f = [X]_k^T \left(\sum_{i=1}^N c_{q_i} c_{q_i}^T \right) [X]_k = [X]_k^T G(q) [X]_k. \tag{2.6}$$

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. In fact, the sum-of-squares representations (with linearly independent q_i's) correspond to factorizations $G = A^T A$ of Gram matrices of f. This motivates the following definition.

Definition 2.1. *We say that a form $f \in \mathbb{R}[X]_{2k}$ is* uniquely sum-of-squares representable *if f has a unique Gram matrix.*

Gram matrices allow one to search for a sum of squares representation of f via the following primal-dual pair of semidefinite programs (SDPs):

$$\begin{array}{lllll}
(2.7) & \text{(Gram)} & & \text{find} & G \succeq 0 \\
 & & & \text{s.t.} & [X]_k^T G [X]_k = f, \\
(2.8) & \text{(Gram)}^* & & \text{minimize} & E(f) \\
 & & & \text{s.t.} & E \in \mathbb{R}[X]_{2k}^\vee \\
 & & & & M_E \succeq 0.
\end{array}$$

Here, for each functional $E\colon \mathbb{R}[X]_{2k} \to \mathbb{R}$, M_E denotes the symmetric bilinear form defined via $M_E(p,q) := E(p \cdot q)$. The convex set

$$\operatorname{Gram}(f) := \{ G \succeq 0 \mid [X]_k^T G [X]_k = f \} \tag{2.9}$$

is the feasible set of the primal SDP. It is called the *Gram spectrahedron* of f. Let us collect some basic properties.

Proposition 2.2. *Let $k \in \mathbb{N}$, $f \in \Sigma_{2k}$.*

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

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

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* $\operatorname{Gram}(f)$ *if there exists some G in* $\operatorname{Gram}(f)$ *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*

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

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 of* $\operatorname{Gram}(f)$ *and this subspace equals* $\operatorname{sosupp} f$.

Proof. Compare the work of Ramana and Goldman [50] on the facial structure of (arbitrary) spectrahedra. Our formulation loosely follows Scheiderer [64, 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 that $x \in \operatorname{relint} F$ if and only if F is the supporting face of x.

2.2. Powers of linear forms. The special 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 [28]). It was later discovered that this uniqueness result can be made algorithmic by interpreting the cubic form as a space of quadratic forms and applying simultaneous matrix diagonalization. The algorithmic result was independently discovered several times, which makes correct attribution nontrivial.

We follow Kolda [36], who argues that the most appropriate single reference is Leurgans, Ross and Abel's 1993 paper; see [41]. However, it should be mentioned that several communities contributed their own variations of the method. Formulations via apolarity theory and catalecticants can be found in [34], [10], a semidefinite programming formulation can be found in [44, section 5.2]. A case can be made that the underlying method was already known to Sylvester [65], albeit his formulation was limited to two variables. It is thus sometimes referred to as *Sylvester's catalecticant method.* The idea has been developed into various directions; see also [15].

In the following, we describe a naive diagonalization method, similar to [41]. The diagonalization method has several numerical issues, but there exist modern, stable alternatives; see, for instance, Anandkumar et al. [4].

Theorem 2.3 (cf. [41]). *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 power sum decomposition*

$$f_d = \sum_{i=1}^{m} \lambda_i \ell_i^d \tag{2.10}$$

such that $\ell_1, \ldots, \ell_m$ *are linearly independent and* $\lambda_1, \ldots, \lambda_m \in \mathbb{R} \setminus \{0\}$, *then compute the forms and weights* $(\ell_1, \lambda_1), \ldots, (\ell_m, \lambda_m)$. *Under these conditions,* (2.10) *is the unique minimum rank power sum decomposition of* (f_2, f_3) *and the only power sum decomposition with linearly independent addends.*

Algorithmic Proof of Theorem 2.3. 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 $\langle \ell_1, \ldots, \ell_m \rangle$, 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 $\ell_i=a_i^Tu$. Compute now the partial derivative

$$f_v := \frac{1}{3}\partial_v f_3 = \sum_{i=1}^m \lambda_i(a_i^Tu(v))(a_i^Tu)^2 = u^TM_vu \tag{2.11}$$

of f_3 in some generic direction $v\in\mathbb{R}^n$. Here, we write $M_v:=\sum_{i=1}^m\lambda_i(a_i^Tu(v))a_ia_i^T$. Similarly, the quadratic form f_2 can be written as $f_2=u^TMu$, where the matrix $M=\sum_{i=1}^m\lambda_ia_ia_i^T\in\mathbb{R}^{m\times m}$ 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

$$\det(M_v-\mu M)=0, \qquad \mu\in\mathbb{R}, \tag{2.12}$$

has m eigenspaces of dimension one, with corresponding eigenvalues $\mu_i:=a_i^Tu(v)$. Indeed, since v was chosen at generically, M_v is of full rank m. The rank of

$$M_v-\mu M=\sum_{i=1}^m\lambda_i((a_i^Tu(v))-\mu)a_ia_i^T \tag{2.13}$$

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

$$M_vx_j=\mu_jMx_j. \tag{2.14}$$

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

$$f_2=\sum_{i=1}^m\nu_i\ell_i^2, \qquad \nu_1,\dots,\nu_m\in\mathbb{R}. \tag{2.15}$$

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

Regarding the uniqueness statement: For any other decomposition $f_d=\sum_{i=1}^{m'}\rho_il_i^d$ 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 $U=\{\partial_vf_2\mid v\in\mathbb{R}^n\}$ must be contained in $\langle l_1,\dots,l_{m'}\rangle$. Since $\dim U=m$, this means that $m'\geq m$. Equality $m=m'$ holds 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. ■

3. An algorithm for power sums. The algorithm from Theorem 2.3 does not extend in any obvious way to higher-order Waring decomposition. However, sometimes it is possible to reduce a k-Waring problem with $k \geq 2$ to a 1-Waring problem, given a small "hint". The hint is a basis of the linear space $\langle q_1, \ldots, q_m \rangle$. Note that for $k = 1$, this subspace can be easily computed, either by taking the space of derivatives of f_2, or the space of second derivatives of f_3. For $k \geq 2$, this will fail since the inner derivative of a form of degree $k \geq 2$ is not a constant.

Therefore, obtaining the space $\langle q_1, \ldots, q_m \rangle$ is a real challenge in the higher-order Waring decomposition, and we argue that it is the main challenge, which separates the $k = 1$ case from the $k \geq 2$ case. Indeed, let us see how the space allows us to reduce to $k = 1$.

Reduction to $k = 1$. Let $u_1, \ldots, u_m$ denote a basis of $\langle q_1, \ldots, q_m \rangle$. Let us grade the algebra $\mathbb{R}[q_1, \ldots, q_m]$ by $\frac{1}{k}$ times the degree. The kernel of the graded algebra homomorphism

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

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

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

admit a joint decomposition as powers of *linear* forms. This means that, given a basis $u_1, \ldots, u_m$ of $\langle q_1, \ldots, q_m \rangle$, it is possible to perform an algorithmic reduction to the case $k = 1$ if the q_i are general enough to not satisfy any algebraic relations of degree 3. 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 $d \leq 3$, there exist unique coefficients $(c_\alpha)_{|\alpha|=d}$ such that

$$f_d = \sum_{|\alpha|=d} c_\alpha u^\alpha, \tag{3.3}$$

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

As an aside, note that this reduction also works if we are just given a basis $u_1, \ldots, u_N$ of a superspace $U \supseteq \langle q_1, \ldots, q_m \rangle$, as long as the $u_1, \ldots, u_N$ do not have any algebraic relations of degree 3. We call this property *cubic independence* of U. The precise definition follows.

Definition 3.1. *Let $u_1, \ldots, u_m$ form in $\mathbb{R}[X]_k$ and $U = \langle u_1, \ldots, u_m \rangle$. We call $u_1, \ldots, u_m$* cubically independent *if the threefold products $(u_i u_j u_k)_{i \leq j \leq k}$ are linearly independent. We call the subspace U* cubically independent *if any of its bases is cubically independent.*

Proposition 4.6 collects some equivalent characterizations.

Space recovery. To obtain the space, we make heuristic use of the Gram spectrahedron. By Proposition 2.2, there is a superspace U_F of $\langle q_1, \ldots, q_m \rangle$ associated with every face F of $\text{Gram}(f_2)$ 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 $G(q)$. In principle, one "only" needs to find one cubically independent superspace U_F of $\langle q_1,\dots,q_m\rangle$. However, it is not clear whether the faces containing $G(q)$ are accessible to us from input f_2, f_3. Spectrahedra often have infinitely many faces and picking the correct ones is not so simple.

Fortunately, the largest facial subspace, $U := \operatorname{sosupp} f_2$, is algorithmically accessible to us, and it contains $G(q)$ by definition. Sometimes, the space U might be too big and not satisfy cubic independence. In those cases, our algorithm fails. In other cases, U might even be equal to $\langle q_1,\dots,q_m\rangle$. This is the case whenever $G(q)$ lies in the relative interior of the Gram spectrahedron. The simplest possible example 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 $q_1,\dots,q_m$. 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 4.8. Our focus lies on the uniquely representable case, since it makes the sum-of-squares support easier to analyze.

To justify that our 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. Theoretical results towards this direction are shown in Ssection 4 and a complementary numerical study is conducted in section 5.

3.1. Algorithms. The procedure to recover the power sum decomposition is described in Algorithm 1. The following theorem is a uniqueness result for power sum 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 3.2 and Algorithm 1, and it holds that $U = \langle q_1,\dots,q_m\rangle$. The condition $\dim \mathbb{R}[U]_{3k} = \binom{m+2}{3}$ is then equivalent to $I_{\mathrm{rel}}(q_1,\dots,q_m)_3 = \{0\}$. Note that a tentative implementation of Algorithm 1, with an example Julia notebook, can be found on GitHub. See [13].

Theorem 3.2. *Let $k \in \mathbb{N}$, and let f_2, f_3 be forms of degree $2k$ and $3k$, respectively. Assume that f_2 is a sum of squares and that the space* $\operatorname{sosupp} f_2$ *is cubically independent. Then, f_2 and f_3 have at most one joint power sum 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 power sum decomposition of (f_2, f_3).

Proof. Denote $U := \operatorname{sosupp} f_2$. Assume there are two distinct power sum 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 2.2, it holds that

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

Algorithm 1. Semidefinite algorithm for power sum decomposition.

Input: $k \in \mathbb{N}$ and forms $f_2 \in \mathbb{R}[X]_{2k}, f_3 \in \mathbb{R}[X]_{3k}$ in variables $X = (X_1, \dots, X_n)$.

Assumptions:

1. f_d have a joint power sum 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. The space $U := \operatorname{sosupp} f_2$ is cubically independent.

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

Procedure:

1: **Compute** some $G \in \operatorname{relint} \operatorname{Gram}(f_2)$ with an interior point SDP solver.
2: **Compute** a basis $u = (u_1, \dots, u_N)$ of $\operatorname{im} G$. `# Note that` $\operatorname{im} G = U$.
3: **Define** N variables $Y_1, \dots, Y_N$.
4: **Define** $\varphi_{\leq 3} \colon \mathbb{R}[Y_1, \dots, Y_N]_{\leq 3} \to \mathbb{R}[u_1, \dots, u_N]_{\leq 3}, Y_i \mapsto u_i$, $(i = 1, \dots, N)$.
`# The assumption` $\dim \mathbb{R}[U]_3 = \binom{N+2}{3}$ `guarantees that` $\varphi_{\leq 3}$ `is invertible.`
5: **for** $d = 2, 3$ **do**
6: **Compute** the unique solution $(c_\alpha)_{|\alpha| = d}$ to the linear system

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

7: **Let** $g_d := \sum_{|\alpha|=d} c_\alpha Y^\alpha$. `# Observe that` $g_d = \varphi_{\leq 3}^{-1}(f_d)$.
8: **end for**
9: **Compute** a decomposition $g_d = \sum_{i=1}^m \lambda_i \ell_i^d$ of g_2, g_3 as power sums of linear forms with the algorithm from Theorem 2.3.
10: **return** $\{(\varphi(\ell_1), \lambda_1), \dots, (\varphi(\ell_m), \lambda_m)\}$.

The linearly independent system $q_1, \dots, q_m$ can therefore be extended to a basis u of U. Write $u = (q_1, \dots, q_m, u_{m+1}, \dots, u_N)$. By cubic independence, there are no algebraic relations of degree 3 between 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 power sum 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 $i \in \{1, \dots, m\}$. ∎

The following simplification of Theorem 3.2 is occasionally useful.

Corollary 3.3. *If $\lambda_1,\dots,\lambda_m \in \mathbb{R}_{>0}$ and $q_1,\dots,q_m \in \mathbb{R}[X]_k$ are forms such that $f_2 := \sum_{i=1}^m \lambda_i q_i^2$ is uniquely sum-of-squares representable and $q_1,\dots,q_m$ are cubically independent, then Algorithm 1 computes the unique powers-of-forms decomposition $\{(q_1,\lambda_1),\dots,(q_m,\lambda_m)\}$ from inputs f_2 and $f_3 := \lambda_1 q_1^3 + \cdots + \lambda_m q_m^3$.*

Proof. If $f_2 = \sum_{i=1}^m \lambda_i q_i^2$ is uniquely sum-of-squares representable, then the space $U :=$ sosupp f_2 equals $\langle q_1,\dots,q_m\rangle$ by Proposition 2.2. Since $q_1,\dots,q_m$ do not satisfy any algebraic relations of degree 3, the space $\mathbb{R}[q_1,\dots,q_m]_3$, spanned by the threefold products of the $q_1,\dots,q_m$, has dimension $\binom{m+2}{3}$ and Theorem 3.2 yields the claim. ∎

Example 3.4. As for our example from (1.1), we can compute with an SDP solver that $f_2 = 2X^4 - 2X^3Y + 3X^2Y^2 - 8XY^3 + 5Y^4$ is uniquely sum-of-squares representable. With the SDP solver MOSEK [46], we compute the Gram matrix of f_2 and certify its uniqueness:

$$G = \begin{array}{c c} & \begin{array}{ccc} X^2 & XY & Y^2 \end{array} \\ \begin{array}{c} X^2 \\ XY \\ Y^2 \end{array} & \begin{pmatrix} 2 & -1 & -1 \\ -1 & 5 & -4 \\ -1 & -4 & 5 \end{pmatrix} \end{array} \tag{3.9}$$

All sum-of-squares decompositions of f_2 must thus arise from factorizations of G as $G = A^TA$. This allows us to find

$$f_2 = \frac{1}{2}(2X^2 - XY - Y^2)^2 + \frac{9}{2}(-XY + Y^2)^2. \tag{3.10}$$

The sum-of-squares support of f_2 hence equals

$$U = \Big\langle \underbrace{2X^2 - XY - Y^2}_{u_1}, \underbrace{-XY + Y^2}_{u_2} \Big\rangle. \tag{3.11}$$

The next step in the algorithm is to express f_2, f_3 in the basis of 3-fold products of $u = (u_1,u_2)$. Recall that $f_3 = -9X^5Y + 18X^4Y^2 - 9X^3Y^3 + 9X^2Y^4 - 18XY^5 + 9Y^6$. We obtain

$$f_3 = \frac{1}{4}\left(9u_1^2u_2 + 27u_2^3\right), \tag{3.12}$$

$$f_2 = \frac{1}{2}\left(u_1^2 + 9u_2^2\right). \tag{3.13}$$

This representation is unique, since u_1, u_2 are cubically independent. By treating u_1, u_2 as variables, we can pretend that f_3 is a cubic form in (u_1,u_2).[2] A 1-Waring decomposition of this cubic form is produced by the algorithm from Theorem 2.3. We obtain

$$f_3 = \left(\frac{1}{2}u_1 + \frac{3}{2}u_2\right)^3 + \left(-\frac{1}{2}u_1 + \frac{3}{2}u_2\right)^3, \tag{3.14}$$

$$f_2 = \left(\frac{1}{2}u_1 + \frac{3}{2}u_2\right)^2 + \left(-\frac{1}{2}u_1 + \frac{3}{2}u_2\right)^2. \tag{3.15}$$

[2]This cubic form is called g_3 in the algorithm.

Now, substituting back yields $q_1 = \frac{1}{2}(u_1 + 3u_2) = X^2 - 2XY + Y^2$ and $q_2 = \frac{1}{2}(3u_2 - u_1) = -X^2 - XY + 2Y^2$.

All the steps in Example 3.4 were calculated with the Julia implementation on GitHub [13]. Note that for the sake of exposition, we rounded errors arising from floating point arithmetic and we normalized the basis u differently, so that it has integer coefficients. By default, our algorithm normalizes the entries of u such that their coefficient vectors have unit length, leading to $u_1 = \frac{1}{\sqrt{6}}(2X^2 - XY - Y^2)$ and $u_2 = \frac{1}{\sqrt{2}}(-XY + Y^2)$. Accounting for that, all steps correspond exactly to outputs of the algorithm provided in [13].

4. Geometric criteria for uniqueness. Let us try to understand the conditions of Corollary 3.3 in terms of geometric properties of the $(q_1, \ldots, q_m) \in \mathbb{R}[X]_k^m$. The following result is a specialization of Theorem 3.2. It gives geometric sufficient success criteria for Algorithm 1.

Corollary 4.1. *Let $q_1, \ldots, q_m \in \mathbb{R}[X]_k^m$ be general k-forms, satisfying one of these properties:*
(1) $m \leq n-2$ and $V_{\mathbb{R}}(q_1, \ldots, q_m)$ contains a nonzero point, or
(2) $m = n-1$ and all lines in the affine cone $V(q_1, \ldots, q_m)$ are real.
Then, $\sum_{i=1}^m q_i^2$ is uniquely sum-of-squares representable. In particular, Algorithm 1 recovers $\{q_1, \ldots, q_m\}$ 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) and (b), respectively, are satisfied on a Euclidean open subset of $\mathbb{R}[X]_k^m$.

The goal of this section is to prove Corollary 4.1. We will need Bertini's theorem and the Poincaré–Miranda theorem. Both will be stated later, in Proposition 4.3 and Theorem 4.5, respectively.

Reminder 4.2. *A subvariety $V \subseteq \mathbb{C}^n$ has (Zariski) 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 4.3 (variant of Bertini's theorem; cf. [29, Chapter II, Exercise 8.4(b), p. 188]). *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 4.4. *Let $m, k \in \mathbb{N}$. Let $q_1, \ldots, q_m \in \mathbb{R}[X]_k$. Assume that $I := (q_1, \ldots, q_m)$ 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* $\operatorname{sosupp} f_2 = \langle q_1, \ldots, q_m \rangle$.

Proof. Let $N \in \mathbb{N}_0$ and $p_1, \ldots, p_N \in \mathbb{R}[X]_k$ 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 $p_1, \ldots, p_m$ vanish on $V_{\mathbb{R}}(I)$. Since $V(I)$ has dense real points and I is radical, $p_1, \ldots, p_m$ must lie in $I_k = \langle q_1, \ldots, q_m \rangle$. Thus $\langle q_1, \ldots, q_m \rangle$ equals $\operatorname{sosupp} f_2$. Claim (b) follows from (a) using Proposition 4.8. ∎

Theorem 4.5 (Poincaré–Miranda, cf. [22, Introduction]). *Write $\mathcal{H}$ for the parallelepiped spanned by linearly independent vectors $v_1, \ldots, v_n \in \mathbb{R}^n$, and let $f \colon \mathcal{H} \to \mathbb{R}^n$ be a continuous function, with coordinates $f(x) = (f_1(x), \ldots, f_n(x))$. Denote*

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

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 are now ready to conclude this section with the proof of Corollary 4.1.

Proof of Corollary 4.1. *Case* 1. Since $m \leq n-2$, Bertini's Proposition 4.3 guarantees that $I = (q_1, \ldots, q_m)$ is a prime ideal and $V(I)$ is smooth and irreducible. Thus, by Reminder 4.2, the existence of a point $x \in V_{\mathbb{R}}(I)$ guarantees that the real points of $V(I)$ are dense. Hence, the condition of Lemma 4.4 is satisfied. The Poincaré–Miranda theorem 4.5 allows one to construct an explicit neighborhood $\mathcal{U}$ such that $V(q_1, \ldots, q_m)$ contains a real nonzero point for all $(q_1, \ldots, q_m) \in \mathcal{U}$: Indeed, let us choose $q_n := 0$ and $q_i = (X_i + X_n)^2 - 2X_n^2 \in \mathbb{R}[X_1, \ldots, X_n, Y]$, for $i \in \{1, \ldots, n-1\}$, and consider the rectangle $\mathcal{H} = [0,1]^n \times \{1\}$ in the affine plane where "$X_n = 1$". Note that $\mathcal{H}_i^a = [0,1]^{i-1} \times \{a\} \times [0,1]^{n-i} \times \{1\}$ 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, in a neighborhood $\mathcal{U}$ of this choice of q_i, the condition of Poincaré–Miranda is satisfied and $V_{\mathbb{R}}(q_1, \ldots, q_m) \neq \emptyset$ is a typical property.

Case 2. Since $m = n-1$, Bertini's theorem yields that $I = (q_1, \ldots, q_m)$ is a radical ideal and $V(I)$ is a union of finitely many points. Thus, the condition of Lemma 4.4 is met if and only if all those lines are real. If $V(I)$ has the maximum number of k^m real points given by the Bézout bound for some specific choice of $q_1, \ldots, q_m$, then so it does in a neighborhood: Indeed, if there was a curve $q(t)$ through $q(0) = (q_1, \ldots, q_m)$ such that $V(q(t))$ had nonreal points for each $t \neq 0$, then a single real point of $V(q(0))$ would have to branch into two complex conjugate points. 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. ■

4.1. Remarks on unique sum-of-squares representability and cubic independence.

Corollary 3.3 has two conditions: cubic independence of $q_1, \ldots, q_m$ and unique representability of $q_1^2 + \cdots + q_m^2$. It is interesting to understand for which values of m these can be satisfied.

Proposition 4.6. *Let* $u_1, \ldots, u_m$ *form in* $\mathbb{R}[X]_k$*, and let* $U = \langle u_1, \ldots, u_m \rangle$*. The following are equivalent characterizations of cubic independence:*

1. *There are no algebraic relations of* $u_1, \ldots, u_m$ *of degree at most* 3.
2. *There are no homogeneous algebraic relations of* $u_1, \ldots, u_m$ *of degree* 3.
3. $\dim \mathbb{R}[u_1, \ldots, u_m]_{3k} = \binom{m+2}{3}$.
4. *The threefold products* $(u_i u_j u_k)_{i \leq j \leq k}$ *are linearly independent.*

Note that for the subspace U*, cubic independence does not depend on the choice of basis.*

As an aside, note that by Proposition 4.6, cubic independence implies linear independence. If $q_1, \ldots, q_m \in U \subseteq \mathbb{R}[X]_k$ are linearly independent, then the tensor $\sum_{i=1}^m q_i^{\otimes 3}$, which is a symmetric 3-tensor on the space U, has a unique (minimum) rank-m decomposition, by Jennrich's uniqueness criterion; see Theorem 2.3. If U is cubically independent, then the projection map $S^3(U) \to \mathbb{R}[X]_{3k}, \sum_{i=1}^m q_i^{\otimes 3} \mapsto \sum_{i=1}^m q_i^3$ is injective. Here, $S^3(U)$ denotes the space of symmetric 3-tensors on U.

Definition 4.7. *We denote by* $\beta_k(n)$ *the maximum dimension of a cubically independent subspace of* $\mathbb{R}[X]_k$*.*

Note that if $m \le \beta_k(n)$, then generic choices of $q_1,\dots,q_m$ will be cubically independent. Clearly, $\beta_k(n) \ge n$, since n generic forms will not have any relations at all. In [7, section 6.4], the authors give a combinatorial proof that $\beta_2(n) = \Theta(n^2)$.

Proposition 4.8. *Let $k, m \in \mathbb{N}$ with $m \le \beta_k(n)$, and let $q_1,\dots,q_m \in \mathbb{R}[X]_k$ 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* $\operatorname{sosupp} f_2 = \langle q_1,\dots,q_m\rangle$.

Proof. If f_2 is uniquely sum-of-squares representable, then clearly $\operatorname{sosupp} f_2 = \langle q_1,\dots,q_m\rangle$. For the converse direction, assume $\operatorname{sosupp} f_2 = \langle q_1,\dots,q_m\rangle$ and let $f_2 = p_1^2 + \dots + p_N^2$ another sum-of-squares representation. Then, $p_1,\dots,p_N$ lie in $\langle q_1,\dots,q_m\rangle$. Write $p_i = c_i^T q = c_{i1}q_1 + \dots + c_{im}q_m$. Then, $f_2 = q^T(\sum_{i=1}^N c_i c_i^T)q = \sum_{i,j=1}^m G_{ij} q_i q_j$, with $G = \sum_{i=1}^N c_i c_i^T \in \mathbb{R}^{m\times m}$. Therefore, $0 = \sum_{i=1}^m q_i^2 - \sum_{i,j=1}^m G_{ij} q_i q_j$, which is a quadratic relation of the q_i's. Since there are no nontrivial quadratic relations, we conclude that$\sum_{i=1}^N c_i c_i^T = G = I_m$. Note that any Gram matrix of f_2 corresponds to a psd $m \times m$ matrix such as G. Therefore, we showed that f_2 is uniquely sum-of-squares representable. ■

Note that the proof of Proposition 4.8 only uses "quadratic independence" of the q_i's.

5. Numerical experiments with trace-free quadratics.

This section conducts a numerical study on random quadratics. In section 4, we saw that uniquely representable sums of squares occur for typical choices of addends if $m \in \{1,\dots,n-1\}$. They likely occur for much larger values of m. In this section, we conduct a numerical study to get a sense both of the possible values of m and of the size of such typical sets, for concrete values of m and n.

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 SDP to determine the dimension of $\operatorname{sosupp}(\sum_{i=1}^m q_i^2)$. The quadratics are chosen from a trace-free distribution, 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 regularly. Precisely, we consider the following distributions.

Definition 5.1. *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 $q := X^T(A - \operatorname{tr}(A)I_n)X$.*

(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 5.4. The code and results of all experiments can be found on GitHub; see [13]. Computations were done in `Julia` [11], with the packages `JuMP` [21], `MultivariatePolynomials` [39], `SumOfSquares` [40], [70], and the `Mosek` solver [46]. For the Gaussian trace-free distribution, this is the observed behavior 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:

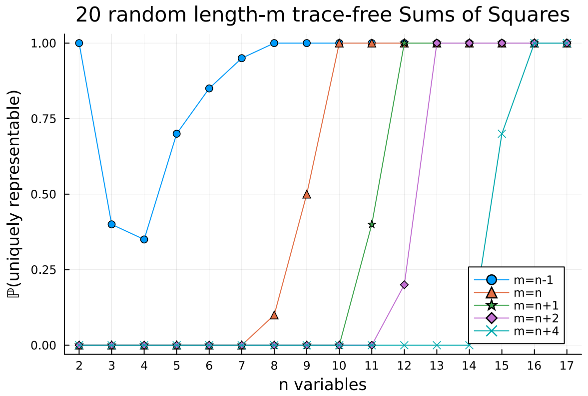

Figure 1. *The y-axis shows the probability that $\sum_{i=1}^m q_i^2$ is uniquely representable if $q_1,\dots,q_m$ are i.i.d. Gaussian random trace-free quadratics. All probabilities were empirically estimated by sampling the quadratic forms and then solving SDPs. Each marker corresponds to an average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

(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 behavior is depicted in Figure 1. Compare [13, `data/experiment-3`].
(2) $p_{m,n}>0$ if $m\le n-1$. See the blue curve in Figure 1. This is consistent with the results from section 4. Compare [13, `data/experiment-3`].
(3) For $m(n)=\lceil\frac{(n+2)(n+1)}{6}\rceil=\Theta(n^2)$, it appears to hold that $p_{m(n),2n}\approx 1$, but $p_{m(n),n}\approx 0$. Cf. [13, `data/experiment-2`].

The same qualitative behavior holds true if the Gaussian trace-free distribution is replaced by the Gauss–Gramian trace-free distribution from Definition 5.1(b). This can be seen from [13, `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 5.2. *If $q_1,\dots,q_n$ are chosen as i.i.d. 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 5.4, for which we gathered separate numerical evidence. In particular, both suggest that there exist open neighborhoods of parameters $q=(q_1,\dots,q_m)$, 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?

Question 5.3. *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) neighborhood of $q_1,\dots,q_{m(n)}$, $\sum_{i=1}^{m(n)} p_i^2$ is uniquely representable?*

Conjecture 5.4. *The explicit family of $m=\lceil\frac{(n+2)(n+1)}{6}\rceil$ trace-free quadratics*

$$q_{ijk}:=(X_i+X_j+X_k)(Y_i+Y_j+Y_k),\qquad (i\le j\le k, i+j+k\equiv 0 \mod n) \tag{5.1}$$

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

$$\sum_{\substack{i\le j\le k\\ i+j+k\equiv 0 \mod n}} q_{ijk}^2 \tag{5.2}$$

is uniquely sum-of-squares representable. 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 $\{X_iY_j \mid i,j=1,\dots,n\}$.

Evidence. Unique sum-of-squares representability was checked numerically for $n=2$, $\dots,15$ in the GitHub [13, `experiments/experiment-1-sos.jl`]. The data files are in `data/experiment-1`. Cubic independence of the q_{ijk}'s was verified with another Julia script, which can be found in `experiments/experiment-1-cubic-independence.jl`. Note that the counting of triplets $i\le j\le k$ such that $i+j+k\equiv 0 \mod n$ is part of the conjecture. The online encyclopedia of integer sequences [47, OEIS sequence A007997] suggests that $m=\lceil\frac{(n+2)(n+1)}{6}\rceil$. ■

6. Applications.

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

6.1. 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 $\lambda_i\ge 0$ being the probability to draw Y_i, assuming $\sum_{i=1}^m \lambda_i = 1$), and then sample Y_i. We denote $Y=\lambda_1Y_1\oplus\cdots\oplus\lambda_mY_m$ 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 $q_i = X^T\Sigma_iX$. It turns out that from sufficiently many samples of Y, it is possible to compute (noisy versions of) the expressions

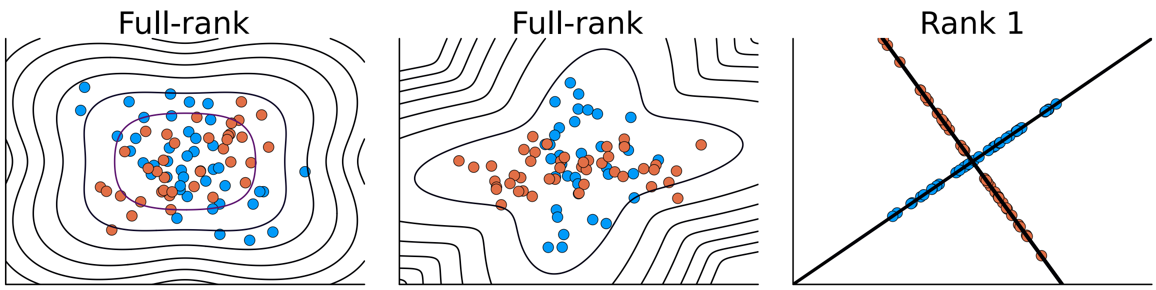

Figure 2. *Three different centered Gaussian mixture distributions of rank 2. The sample coloring indicates which of the Gaussians was chosen in the sampling process. The pictures on the left and middle show mixtures of full-dimensional Gaussians as treated in subsection 6.1. Here, the contour lines describe the probability density function of the mixture. The rightmost picture shows a mixture on proper subspaces. These are addressed in subsection 6.2.*

$$\sum_{i=1}^{m} \lambda_i q_i^d, \qquad d=0,\dots,D, \tag{6.1}$$

where $D \in \mathbb{N}$ is some threshold depending on the order of the number of samples. Up to scalars, the expressions (6.1) 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], and with slightly less detail in [14, section 2].

The goal is to estimate the parameters, that is, the symmetric covariance matrices Σ_i (corresponding to the q_i) as well as the *mixing weights* λ_i 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 $\mathcal{O}(\sigma^d)$ i.i.d. 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 [14] 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 $D=3$. Adapting Theorem 3.2 to the setting of Gaussian mixtures yields the following result.

Proof of Corollary 1.1. Writing $q_i = X^T \Sigma_i X$, the moment forms $\mathcal{M}_{2d}(Y)$ of the Gaussian mixture random variable $Y = \lambda_1 \mathcal{N}(0,\Sigma_1) \oplus \dots \oplus \lambda_m \mathcal{N}(0,\Sigma_m)$ may be expressed as (cf. [12, Chapter 3, Introduction])

$$\mathcal{M}_{2d}(Y) = \frac{1}{d!2^d} \sum_{i=1}^{m} \lambda_i q_i^d, \tag{6.2}$$

and these are given as input for $d \in \{0,\dots,3\}$. By Corollary 4.1, 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

$$f_2 = \sum_{i=1}^{m} \lambda_i (q_i - Z)^2, \quad f_3 = \sum_{i=1}^{m} \lambda_i (q_i - Z)^3 \tag{6.3}$$

from the input. This is possible since the X-homogeneous parts of the power sums from (6.3) correspond to the moments forms $\mathcal{M}_0(Y), \mathcal{M}_2(Y), \mathcal{M}_4(Y), \mathcal{M}_6(Y)$. 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 6.1. The numerical experiments, which lead to Conjecture 5.2, suggest that the shifting method from the proof of Corollary 1.1 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 i.i.d. from the distribution $\frac{1}{n^2}X^TA^TAX$, where A is a random $n\times n$ matrix with i.i.d. 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 $\sigma^2\approx\frac{1}{m}\operatorname{tr}(\sum_{i=1}^m q_i)$. This motivates Definition 5.1(b) of Gauss–Gramian quadratic forms, since that will be the distribution of Gaussian random psd forms after shifting by the trace.

6.2. 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], [30], [42], [32]. 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. For example, 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 subsection 6.1, with the difference that Gaussians on subspaces will lead to psd forms $q_1,\dots,q_m$ that are not of full rank.

Proof of Corollary 1.2. 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 $\lambda_1Y_1\oplus\dots\oplus\lambda_mY_m$, up to known scalars, attain the form

$$\sum_{i=1}^m \lambda_i q_i^d, \qquad d=0,1,2,3. \tag{6.4}$$

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 Theorem 2.3. Thus, w.l.o.g. $r\geq 2$.

Consider first a special instance: For $m\leq n-1$, $I=(X_1^2-X_n^2,\dots,X_m^2-X_n^2)$ 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 [23, 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 neighborhood $\mathcal{U}$ 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 neighborhood 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 neighborhood: 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 nonreal points, then a pair of complex conjugate nonreal 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 neighborhood 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 neighborhood $\mathcal{U}'$ of $q = q_{\text{spec}}$.

It is easy to see that the forms $X_i^2 - X_n^2$ are cubically independent, since each has a different variable. By Proposition 6.2, which we prove later, cubic independence still holds in a neighborhood.

Summarized, this shows by Lemma 4.4 that 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

$$\sum_{i=1}^{m} (q_i - Z)^2 \quad \text{and} \quad \sum_{i=1}^{m} (q_i - Z)^3 \tag{6.5}$$

from the input.

(2) Plug $Z \mapsto 2X_n$ into the forms from (6.3) 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 neighborhood of a point in PSD_r contains points in the interior of PSD_r. This means that $\mathcal{U}' \cap \mathrm{PSD}_r$ 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. ∎

It remains to show the following proposition, which we used in the proof of Corollary 1.2 for the special case $\mathcal{D} = \mathcal{D}_r$.

Proposition 6.2. *Let $m, n, k, d \in \mathbb{N}_0$. Let $\mathcal{D} \subseteq \mathbb{C}[X]_k$ an irreducible variety containing m distinct forms that do not satisfy any relations of degree d. Let $q_1, \dots, q_m$ general in $\mathcal{D}$. Then also $I_{\text{rel}}(q_1, \dots, q_m)_d = \{0\}$.*

Proof. Let $q = (q_1, \dots, q_m) \in \mathcal{D}^m$ such that $I_{\text{rel}}(q_1, \dots, q_m)_d = \{0\}$. Having no algebraic relations is a (Zariski) open condition, so we find an open neighborhood $\mathcal{U} \subseteq K[X]_k^m$ of q such that for all elements of $\mathcal{U}$ the d-fold products of their entries are linearly independent. Thus, this property holds on the dense open subset $\mathcal{U} \cap \mathcal{D}^m$ of $\mathcal{D}^m$. ∎

Acknowledgments. The author wishes to thank Pravesh Kothari, who encouraged this work on powers-of-forms decompositions and pointed me towards the work of Ankit Garg [25]. The author also wish to thank Monique Laurent for the rich feedback provided to this article, in particular for catching several mistakes. The author further wishes 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.

REFERENCES

[1] J. Alexander, and A. Hirschowitz, *Polynomial interpolation in several variables*, J. Algebraic Geom., 4 (1995), pp. 201–222.

[2] C. Amendola, J.-C. Faugere, and B. Sturmfels, *Moment varieties of Gaussian mixtures*, J. Algebr. Stat., 7 (2016), pp. 14–28, https://doi.org/10.18409/jas.v7i1.42.

[3] C. Améndola, K. Ranestad, and B. Sturmfels, *Algebraic identifiability of Gaussian mixtures*, Int. Math. Res. Not., 2018 (2018), pp. 6556–6580, https://doi.org/10.1093/imrn/rnx090.

[4] A. Anandkumar, R. Ge, D. Hsu, S. M. Kakade, and M. Telgarsky, *Tensor decompositions for learning latent variable models*, J. Mach. Learn. Res., 15 (2014), pp. 2773–2832, https://dl.acm.org/doi/10.5555/2627435.2697055.

[5] J. Anderson, M. Belkin, N. Goyal, L. Rademacher, and J. R. Voss, *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.

[6] E. Angelini, F. Galuppi, M. Mella, and G. Ottaviani, *On the number of Waring decompositions for a generic polynomial vector*, J. Pure Appl. Algebra, 222 (2018), pp. 950–965, https://doi.org/10.1016/j.jpaa.2017.05.016.

[7] M. Bafna, T. Hsieh, P. Kothari, and J. Xu, *Polynomial-time power-sum decomposition of polynomials*, in Proceedings of the 2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS), 2022, pp. 956–967.

[8] E. Ballico, A. Bernardi, M. V. Catalisano, and L. Chiantini, *Grassmann secants, identifiability, and linear systems of tensors*, Linear Algebra Appl., 438 (2013), pp. 121–135, https://doi.org/10.1016/j.laa.2012.07.045.

[9] A. S. Bandeira, B. Blum-Smith, J. Kileel, A. Perry, J. Weed, and A. S. Wein, *Estimation under Group Actions: Recovering Orbits From Invariants*, preprint, https://arxiv.org/abs/1712.10163, 2017.

[10] A. Bernardi, A. Gimigliano, and M. Idà, *Computing symmetric rank for symmetric tensors*, J. Symbolic Comput., 46 (2011), pp. 34–53, https://doi.org/10.1016/j.jsc.2010.08.001.

[11] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, *Julia: A fresh approach to numerical computing*, SIAM Rev., 59 (2017), pp. 65–98, https://doi.org/10.1137/141000671.

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

[13] A. T. Blomenhofer, *Code and Files for: Unique Powers-of-Forms Decompositions from Simple Gram Spectrahedra*, 2023, available at github.com/a44l/powers-of-forms.

[14] A. T. Blomenhofer, A. Casarotti, A. Oneto, and M. Michałek, *Identifiability for mixtures of centered Gaussians and sums of powers of quadratics*, Bull. Lond. Math. Soc., 55 (2023), pp. 2407–2424, https://doi.org/10.1112/blms.12871.

[15] J. Brachat, P. Comon, B. Mourrain, and E. Tsigaridas, *Symmetric tensor decomposition*, Linear Algebra Appl., 433 (2010), pp. 1851–1872, https://doi.org/10.1016/j.laa.2010.06.046.

[16] L. Chiantini, G. Ottaviani, and N. Vannieuwenhoven, *On generic identifiability of symmetric tensors of subgeneric rank*, Trans. Amer. Math. Soc., 369 (2016), pp. 4021–4042, https://doi.org/10.1090/tran/6762.

[17] R. E. Curto and P. J. di Dio, *Time-dependent moments from the heat equation and a transport equation*, Int. Math. Res. Not., 17 (2023), pp. 14955–14990, https://doi.org/10.1093/imrn/rnac244.

[18] S. Dasgupta, *Learning mixtures of Gaussians*, in Proceedings of the 40th Annual Symposium on Foundations of Computer Science, FOCS '99, New York, 1999, pp. 634–644.
[19] S. Dasgupta and L. Schulman, *A probabilistic analysis of em for mixtures of separated, spherical Gaussians*, J. Mach. Learn. Res., 8 (2007), pp. 203–226, https://dl.acm.org/doi/pdf/10.5555/1314498.1314505.
[20] P. J. di Dio, *The multidimensional truncated moment problem: Gaussian mixture reconstruction from derivatives of moments*, J. Math. Anal. Appl., 517 (2023), pp. 126592, https://doi.org/10.1016/j.jmaa.2022.126592.
[21] I. Dunning, J. Huchette, and M. Lubin, *JuMP: A modeling language for mathematical optimization*, SIAM Rev., 59 (2017), pp. 295–320, https://doi.org/10.1137/15M1020575.
[22] A. Fonda and P. Gidoni, *Generalizing the Poincaré–Miranda theorem: The avoiding cones condition*, Ann. Mat. Pura Appl., 195 (2015), pp. 1347–1371, https://doi.org/10.1007/s10231-015-0519-6.
[23] W. Fulton, *Intersection Theory*, Ergebnisse der Mathematik und ihrer Grenzgebiete, Springer, New York, 2012.
[24] F. Galuppi and M. Mella, *Identifiability of homogeneous polynomials and Cremona transformations*, Journal für die reine und angewandte Mathematik (Crelles Journal), 2019 (2019), pp. 279–308, https://doi.org/10.1515/crelle-2017-0043.
[25] A. Garg, N. Kayal, and C. Saha, *Learning sums of powers of low-degree polynomials in the nondegenerate case*, in Proceedings of the 2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS), IEEE Computer Soc., Los Alamitos, CA, 2020, pp. 889–899.
[26] R. Ge, Q. Huang, and S. M. Kakade, *Learning mixtures of Gaussians in high dimensions*, in Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing, ACM, New York, 2015, pp. 761–770.
[27] F. Gesmundo, A. Oneto, and E. Ventura, *Partially symmetric variants of Comon's problem via simultaneous rank*, SIAM J. Matrix Anal. Appl., 40 (2019), pp. 1453–1477, https://doi.org/10.1137/18M1225422.
[28] R. Harshman, *Foundations of the parafac procedure: Models and conditions for an "explanatory" multimodal factor analysis*, UCLA Working Papers in Phonetics, 16, (1970), pp. 1–84.
[29] R. Hartshorne, *Algebraic Geometry*, Vol. 52, Springer Science & Business Media, 2013.
[30] C. Hegde, P. Indyk, and L. Schmidt, *Fast recovery from a union of subspaces*, in Adv. Neural Inf. Proc. Sys. NIPS, 2016, pp. 4401–4409, https://dl.acm.org/doi/10.5555/3157382.3157589.
[31] D. Hilbert, *Lettre adressée à M. Hermite*, 1933, pp. 148–153, http://www.numdam.org/item/JMPA_1888_4_4__249_0/.
[32] D. Hong, R. P. Malinas, J. A. Fessler, and L. Balzano, *Learning dictionary-based unions of subspaces for image denoising*, in Proceedings of the 2018 26th European Signal Processing Conference (EUSIPCO), 2018, pp. 1597–1601.
[33] D. J. Hsu and S. M. Kakade, *Learning mixtures of spherical Gaussians: Moment methods and spectral decompositions*, in Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, 2013, pp. 11–20.
[34] A. Iarrobino and V. Kanev, *Power Sums, Gorenstein Algebras, and Determinantal Loci*, Lecture Notes in Math. 1721, Springer-Verlag, Berlin, 1999.
[35] A. T. Kalai, A. Moitra, and G. Valiant, *Efficiently learning mixtures of two Gaussians*, in STOC'10—Proceedings of the 2010 ACM International Symposium on Theory of Computing, ACM, New York, 2010, pp. 553–562.
[36] T. G. Kolda, *Will the real Jennrich's Algorithm please stand up?*, 2021, https://www.mathsci.ai/post/jennrich/, accessed April 6, 2024.
[37] J. M. Landsberg, *Tensors: Geometry and Applications*, American Mathematical Society, Providence, RI, 2012.
[38] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging Applications of Algebraic Geometry, IMA Vol. Math. Appl. 149, Springer, New York, 2009, pp. 157–270, https://doi.org/10.1007/978-0-387-09686-5_7.
[39] B. Legat, *Multivariate polynomials in Julia*, July 2022. in JuliaCon. *Available at:* https://pretalx.com/juliacon-2022/talk/TRFSJY/.

[40] B. Legat, C. Coey, R. Deits, J. Huchette, and A. Perry, *Sum-of-squares optimization in Julia*, in The First Annual JuMP-dev Workshop, 2017.
[41] S. E. Leurgans, R. T. Ross, and R. B. Abel, *A decomposition for three-way arrays*, SIAM J. Matrix Anal. Appl., 14 (1993), pp. 1064–1083, https://doi.org/10.1137/0614071.
[42] J. Lipor and L. Balzano, *Leveraging union of subspace structure to improve constrained clustering*, in Proceedings of the 34th International Conference on Machine Learning, Vol. 70, ICML'17, JMLR.org, 2017, pp. 2130–2139.
[43] A. Liu and A. Moitra, *Settling the robust learnability of mixtures of Gaussians*, in Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, 2021, pp. 518–531, https://dl.acm.org/doi/10.1145/3406325.3451084.
[44] T. Ma, J. Shi, and D. Steurer, *Polynomial-time tensor decompositions with sum-of-squares*, in FOCS, I. Dinur, ed., IEEE Computer Society, 2016, pp. 438–446.
[45] A. Moitra and G. Valiant, *Settling the polynomial learnability of mixtures of Gaussians*, in Proceedings of the 2010 IEEE 51st Annual Symposium on Foundations of Computer Science, 2010, pp. 93–102, https://doi.org/10.1109/FOCS.2010.15.
[46] MOSEK-ApS, *Semidefinite Optimization — Mosek Optimizer API for Python*, 2019.
[47] OEIS Foundation Inc., *Entry A007997 in The On-Line Encyclopedia of Integer Sequences*, 2024, https://oeis.org/A007997, accessed May 5, 2024.
[48] K. Pearson, *Mathematical contributions to the theory of evolution*. VII. *On the correlation of characters not quantitatively measurable*, Philos. Trans. R. Soc. Lond. Ser. A, Containing Papers of a Mathematical or Physical Character, 195 (1900), pp. 1–47+405.
[49] H. H. Permuter, J. M. Francos, and I. H. Jermyn, *Gaussian mixture models of texture and colour for image database retrieval*, in Proceedings of the 2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP '03, Hong Kong, 2003, pp. 569–572.
[50] M. Ramana and A. J. Goldman, *Some geometric results in semidefinite programming*, J. Global Optim., 7 (1995), pp. 33–50, https://doi.org/10.1007/BF01100204.
[51] S. Ramanujan, *Problem* 441, J. Indian Math. Soc., V (1913), 29.
[52] S. Ramanujan, *Collected Papers of Srinivasa Ramanujan*, AMS Chelsea, Providence, RI, 2000.
[53] O. Regev and A. Vijayaraghavan, *On learning mixtures of well-separated Gaussians*, in Proceedings of the 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2017, pp. 85–96.
[54] D. A. Reynolds and R. C. Rose, *Robust text-independent speaker identification using Gaussian mixture speaker models*, IEEE Trans. Speech Audio Process., 3 (1995), pp. 72–83, https://doi.org/10.1109/89.365379.
[55] B. Reznick, *Sums of even powers of real linear forms*, Mem. Amer. Math. Soc., 96 (1992), 463, https://doi.org/10.1090/memo/0463.
[56] B. Reznick, *Patterns of dependence among powers of polynomials*, in Algorithmic and Quantitative Real Algebraic Geometry (Piscataway, NJ, 2001), DIMACS Ser. Discrete Math. Theoret. Comput. Sci. 60, Amer. Math. Soc., Providence, RI, 2003, pp. 101–121.
[57] B. Reznick, *Laws of inertia in higher degree binary forms*, Proc. Amer. Math. Soc., 138 (2010), pp. 815–826, https://doi.org/10.1090/S0002-9939-09-10186-7.
[58] B. Reznick, *On the length of binary forms*, in Quadratic and Higher Degree Forms, Dev. Math. 31, Springer, New York, 2013, pp. 207–232.
[59] B. Reznick, *Some new canonical forms for polynomials*, Pacific J. Math., 266 (2013), pp. 185–220, https://doi.org/10.2140/pjm.2013.266.185.
[60] B. Reznick, *Linearly dependent powers of binary quadratic forms*, Pacific J. Math., 303 (2019), pp. 729–755, https://doi.org/10.2140/pjm.2019.303.729.
[61] B. Reznick, *Equal sums of two cubes of quadratic forms*, Int. J. Number Theory, 17 (2021), pp. 761–786, https://doi.org/10.1142/S1793042120400308.
[62] B. Reznick and N. Tokcan, *Binary forms with three different relative ranks*, Proc. Amer. Math. Soc., 145 (2017), pp. 5169–5177, https://doi.org/10.1090/proc/13666.
[63] A. Sanjeev and R. Kannan, *Learning mixtures of arbitrary Gaussians*, in Proceedings of the Thirty-Third Annual ACM Symposium on Theory of Computing, 2001, pp. 247–257.

[64] C. Scheiderer, *Extreme points of Gram spectrahedra of binary forms*, Discrete Comput. Geom., 67 (2022), pp. 1174–1190, https://doi.org/10.1007/s00454-022-00385-w.

[65] J. J. Sylvester, *Sur une extension d'un théorème de Clebsh relatif aux courbes du quatrième degré*, C. R. Math. Acad. Sci. Paris, 102 (1886), pp. 1532–1534.

[66] J. J. Sylvester, *Collected Works*, Cambridge University Press, Cambridge, 1904.

[67] A. Terracini, *Sulle V_k che rappresentano più di$\frac{k(k-1)}{2}$ equazioni di laplace linearmente indipendenti*, Rend. Circ. Mat. Palermo, 33 (1912), pp. 176–186.

[68] A. Terracini, *Sulla rappresentazione delle coppie di forme ternarie mediante somme di potenze di forme lineari*, Ann. Mat. Pura Appl. (1898-1922), 24 (1915), pp. 1–10, https://doi.org/10.1007/BF02419670.

[69] Y. Wang, D. Wipf, Q. Ling, W. Chen, and I. Wassell, *Multi-task learning for subspace segmentation*, in Proceedings of the 32nd International Conference on Machine Learning, Proceedings of Machine Learning Research 37, F. Bach and D. Blei, eds., PMLR, Lille, France, 2015, pp. 1209–1217.

[70] T. Weisser, B. Legat, C. Coey, L. Kapelevich, and J. P. Vielma, *Polynomial and moment optimization in Julia and JuMP*, in *JuliaCon*, 2019.