

IDENTIFIABILITY FOR MIXTURES OF CENTERED GAUSSIANS AND SUMS OF POWERS OF QUADRATICS

ALEXANDER TAVEIRA BLOMENHOFER, ALEX CASAROTTI, MATEUSZ MICHAŁEK,
AND ALESSANDRO ONETO

ABSTRACT. We consider the inverse problem for the polynomial map which sends an m -tuple of quadratic forms in n variables to the sum of their d -th powers. This map captures the moment problem for *mixtures of m centered n -variate Gaussians*. In the first non-trivial case $d = 3$, we show that for any $n \in \mathbb{N}$, this map is generically one-to-one (up to permutations of q_1, \dots, q_m and third roots of unity) in two ranges: $m \leq \binom{n}{2} + 1$ for $n \leq 16$ and $m \leq \binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1$ for $n > 16$, thus proving *generic identifiability* for mixtures of centered Gaussians from their (exact) moments of degree at most 6. The first result is obtained by the explicit geometry of the tangential contact locus of the variety of sums of cubes of quadratic forms, as described in [9], while the second result is accomplished using the link between secant non-defectivity with identifiability, proved in [6]. The latter approach generalizes also to sums of d -th powers of k -forms for $d \geq 3$ and $k \geq 2$.

1. INTRODUCTION

1.A. **Motivation.** Ever since the pioneering work of Pearson (original work in [26], but see e.g. [1] for a modern exposition) on the separation of biological species, *Gaussian mixture models* are a highly important tool for modern data analysis. Given the moments of a mixture Y of n -dimensional normally distributed random variables $Y_i \sim \mathcal{N}(\mu_i, \Sigma_i)$, one aims to recover the parameters $(\mu_1, \Sigma_1), \dots, (\mu_m, \Sigma_m)$ (up to permutation). One of the first and fundamental questions to ask is under which circumstances there exists a unique solution, essentially justifying that the parameters bear information which is “meaningful” for statistical inference. In this work, we are concerned with the special case of centered Gaussians, i.e., $\mu_1 = \dots = \mu_m = 0 \in \mathbb{R}^n$, and we focus on moments of one, fixed degree. Then, the estimation problem turns out to be profoundly related to an algebraic problem. Identify $S^d(\mathbb{C}^n)$ with the vector space of degree- d homogeneous polynomials in n -variables with complex coefficients. Given a sum

$$(1.1) \quad \sum_{i=1}^m q_i^d \in S^{2d}(\mathbb{C}^n)$$

of powers of quadratic forms $q_1, \dots, q_m \in S^2(\mathbb{C}^n)$, when can we obtain the addends q_1^d, \dots, q_m^d up to permutation? This is the *decomposition problem for powers of quadratic forms*. Note that if the q_i 's are the quadratic forms corresponding to the covariance matrices Σ_i 's of the centered Gaussian random vectors Y_i 's, then, up to a factor of $\frac{1}{m}$, (1.1) corresponds to the $2d$ -th moment of their (uniformly weighted) mixture Y . We will fully explain the connection in Section 2. While of course the decomposition problem asks to *obtain* the addends in the sense of algorithmic computation, our set objective in the present work is just to examine when there exists a unique solution. In that case, we say that *identifiability* holds. The smallest nontrivial case (and arguably also the most interesting one) is when moments of degree 6 are given, that is $d = 3$. Indeed, note that for $d = 2$ we cannot hope for identifiability since $q_1^2 + q_2^2 = \frac{1}{2}(q_1 + q_2)^2 + \frac{1}{2}(q_1 - q_2)^2$.

2020 *Mathematics Subject Classification.* primary: 15A69, 14Q20 .

MM is funded by the Deutsche Forschungsgemeinschaft — Projektnummer 467575307.

AO is partially supported by MIUR and GNSAGA of INdAM (Italy).

1.B. Statement of main results. In the case of sextics as sums of cubes of quadratic forms, we obtain the following result about general identifiability.

Theorem 1.1. *Let $n, m \in \mathbb{N}$ such that one of the following holds:*

- (1) $n > 16$ and $m \leq \binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1$; or
- (2) $n \leq 16$ and $m \leq \binom{n}{2} + 1$.

Then, for general $q_1, \dots, q_m \in S^2(\mathbb{C}^n)$, the sextic $t = \sum_{i=1}^m q_i^3$ has a unique representation as a sum of m cubes of quadratic forms, up to permutation and third roots of unity.

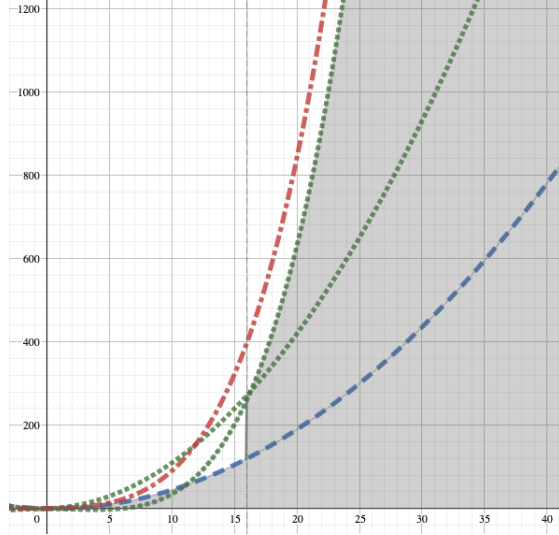


FIGURE 1. The ranges covered by the main result (Theorem 1.1) while letting the number of variables n increase along the horizontal axis. The green-dotted curves refer to condition (1): note that the result holds only for $n > 16$ because it requires that $\binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1 > 2 \left(\binom{n+1}{2} - 1 \right)$ (cfr. Corollary 4.4). The blue-dashed line refers to condition (2) (cfr. Section 4.B). The red-dashed-dotted line is the expected number of cubes needed to write a general complex sextic as their sum, i.e., $\lfloor \dim S^6(\mathbb{C}^n) / \dim S^2(\mathbb{C}^n) \rfloor$: recall that if there are more addends than that number generic identifiability is impossible by dimensionality.

With the connection explained in Section 2, we directly obtain the following consequence about degree-6 moments of mixtures of general centered Gaussians.

Corollary 1.2. *Let $n, m \in \mathbb{N}$ such that one of the following holds:*

- (1) $n > 16$ and $m \leq \binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1$; or
- (2) $n \leq 16$ and $m \leq \binom{n}{2} + 1$.

Let $Y_1 \sim \mathcal{N}(0, q_1), \dots, Y_m \sim \mathcal{N}(0, q_m)$ centered normal distributions given by general psd covariance forms $q_1, \dots, q_m \in S^2(\mathbb{R}^n)$. Let Z_1, \dots, Z_m be any other centered Gaussian random vectors on \mathbb{R}^n such that one of the following holds:

- (a) *the uniformly weighted Gaussian mixtures $Y = \frac{1}{m}(Y_1 \oplus \dots \oplus Y_m)$ and $Z = \frac{1}{m}(Z_1 \oplus \dots \oplus Z_m)$ agree on the moments of degree 6; or*
- (b) *for general $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{>0}$ and for $\mu_1, \dots, \mu_m \in \mathbb{R}_{\geq 0}$ both summing up to 1, the Gaussian mixtures $Y = \lambda_1 Y_1 \oplus \dots \oplus \lambda_m Y_m$ and $Z = \mu_1 Z_1 \oplus \dots \oplus \mu_m Z_m$ agree on the moments of degree 6 and 4.*

Then $\{Y_1, \dots, Y_m\} = \{Z_1, \dots, Z_m\}$ and $Y = Z$. In case (b), the corresponding mixing weights are equal, too.

Condition (1) of Theorem 1.1 may be generalized to guarantee generic identifiability for sums of arbitrary powers of high degree forms (cfr. Corollary 4.4). We highlight it here in the case of powers of quadratic forms, which is the case that is relevant for mixtures of centered Gaussians.

Theorem 1.3. *Fix $d \in \mathbb{N}$. For any $n, m \in \mathbb{N}$ such that*

$$(1.2) \quad 3 \binom{n+1}{2}^2 - 2 \binom{n+1}{2} < \binom{n-1+2d}{2d}$$

and

$$m \leq \binom{n-1+2d}{2d} / \binom{n+1}{2} - \binom{n+1}{2} - 1$$

Then, for general $q_1, \dots, q_m \in S^2(\mathbb{C}^n)$, the degree- $2d$ form $t = \sum_{i=1}^m q_i^d$ has a unique representation as a sum of m d -th powers of quadratic forms, up to permutation and third roots of unity.

With the connection explained in Section 2, we obtain an analogous of Corollary 1.2 for moments of degree $2d$.

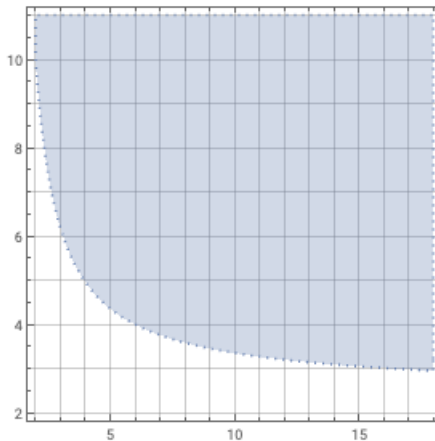


FIGURE 2. The region of pairs (n, d) satisfying inequality (1.2) and for which Theorem 1.3 holds.

1.C. Outline of this paper. In Section 2, we give a concise explanation of the connection between the decomposition problem for cubes of quadratics and the moment problem for mixtures of centered Gaussians. In Section 3 we recall the basic facts on the theory of secant varieties while in Section 4 we recall the general identifiability results that are then used to prove the main results. Some parts of our analysis are verified on a computer. The computation is done with the `Julia` programming language [4] and the `MultivariatePolynomials.jl` library [21]. Our code can be found on GitHub, see [28].

1.D. Related work. On the algorithmic side, Ge, Huang and Kakade provided an algorithm that learns the parameters of mixtures of n -variate Gaussians of rank at most $\mathcal{O}(\sqrt{n})$ from their moments of degree at most 6 (see [16]). The paper considers a smoothed analysis framework, which is essentially a numerically stable way of saying that the quadratics should be in general position. Mixtures of *centered* Gaussians were recently studied in [15], with a particular focus towards their relation to the decomposition problem for sums of powers of quadratics as well as circuit complexity. The work of Garg, Kayal and Saha [15] also studies the more general circuit model of sums of powers of low degree polynomials and was in fact a major motivation for the present work. Sums of powers have been studied from

the point of view of algebraic geometry, e.g. [13, 22], and recently appeared also in relation to polynomial neural networks [20].

In the usual terminology of additive decompositions, fixed positive integers d and k , the (k, d) -rank (or short *rank*) of a degree- dk form is the smallest number of degree- k forms needed to write the given form as their sum of their d -th powers. From an algebro-geometric point of view, additive decompositions of polynomials are studied through secant varieties. In our context, the m -th secant variety is the (Zariski) closure of the set of polynomials of degree dk of rank at most m .

Several interesting open questions are yet to be answered for decompositions of homogeneous polynomials as sums of powers. First of all, what is the general rank, i.e., the smallest number of d -th powers needed to write a general form of degree dk as their sum? For $k = 1$, i.e., sums of powers of linear forms, it is worth recalling that the question is answered by the celebrated Alexander-Hirschowitz Theorem [19] (see also [5]). For $k \geq 2$, the complete answer is given in the cases of binary forms and sums of squares in three and four variables in [22]. In general, it is conjectured that the generic rank is as expected by dimension count $\left\lceil \binom{n+dk-1}{dk} / \binom{n+k-1}{k} \right\rceil$ unless $d = 2$ (cfr. [22, Conjecture 1.2]). The latter question would be answered by knowing dimensions of all secant varieties.

As for general identifiability, i.e., uniqueness of the decomposition for a general point of a secant variety, a complete answer is given in the case of sums of powers of linear forms ($k = 1$): In [11] it was shown that in all but a few exceptional cases identifiability holds for all subgeneric ranks, while Galuppi [14] completed the classification of cases in which identifiability holds also for generic ranks. For $k \geq 2$, as far as we know, before the present work, only identifiability for sextics as sums of cubes was recently addressed for rank 2 [27].

Generic identifiability in the range (1) of Theorem 1.1 is proved employing a result by Casarotti and Mella [6], which translates the study of general identifiability to the study of dimensions of secant varieties under certain constraints on rank and dimension. The dimension of the secant varieties of varieties of powers is given by the main result of [24]. In the range (2) of Theorem 1.1, our analysis employs the geometric notion of weak defectivity and tangential contact loci due to Chiantini and Ciliberto [7, 8] and is based on a series of works due to Chiantini, Ottaviani and Vannieuwenhoven [9, 10, 11], where the authors examine the question of *generic tensor identifiability*, i.e., under which conditions does a general tensor of fixed rank m have a unique decomposition as a sum of m simple (rank-1) tensors. Symmetric tensor decomposition corresponds to the decomposition problem for powers of *linear* forms.

Specifically for cubes of linear forms in n variables, the generic rank is in $\theta(n^2)$, whereas efficient algorithms for the decomposition problem succeeding in the smoothed analysis framework are known up to rank n (e.g. [2], with the original idea dating back to Jennrich, published via Harshman [18]). Various other algorithms for symmetric tensor decomposition of order 3 exist. Some efficient algorithms can exceed the rank- n threshold for “average case” problems and go up to rank almost $n^{1.5}$ (e.g. [17], [23]), by relying on the assumption that the rank-1 components are drawn from a friendly distribution. Other algorithms can produce decompositions for all subgeneric ranks, but sacrifice computational efficiency, cf. e.g. the work of Bernardi and Taufer [3]. This leaves a multiplicative gap of $\theta(n)$ between the regime where generic identifiability holds and the regime where the rank-1 components can be efficiently computed in a smoothed analysis framework.

For quadratic forms, the threshold of $\mathcal{O}(\sqrt{n})$ due to Ge, Huang and Kakade [16] might not be the final answer either. One might conjecture that efficient algorithms are possible at least as long as the rank is at most the number of variables. Any algorithm succeeding for superquadratic rank $m \gg \dim S^2(\mathbb{C}^n)$ would by the aforementioned also have nontrivial implications on tensor decomposition, as clearly from a 3-tensor $\sum_{i=1}^m q_i^{\otimes 3} \in S^3(S^2(\mathbb{C}^n))$ we may compute the 6-form $\sum_{i=1}^m q_i^3$ by applying a linear map. This explains our focus towards finding specific witnesses for quadratic-rank generic identifiability in Section 4.B, although we stress that our results do *not* have algorithmic consequences.

ACKNOWLEDGEMENTS

We thank Pravesh Kothari for giving inspiring questions and encouraging comments as well as pointing us to the work of Garg, Kayal and Saha [15]. We thank Nick Vannieuwenhoven for helpful explanations and both Boris Shapiro and an anonymous reader for telling us about additional, very interesting references. We thank Joseph Landsberg and Laurent Manivel for organizing a "Workshop on geometry and complexity theory" where two of the authors had the opportunity to exchange ideas.

2. SUMS OF POWERS OF QUADRATICS AND MIXTURES OF CENTERED GAUSSIANS

Notation. For any field \mathbb{F} , let $X = (X_1, \dots, X_n)$ be a set of variables of \mathbb{F}^n . Let $\mathbb{F}[X] = \mathbb{F}[X_1, \dots, X_n] = \bigoplus_{d \geq 0} \mathbb{F}[X]_d$ be the standard graded polynomial ring where $\mathbb{F}[X]_d$ is identified with $S^d(\mathbb{F}^n)$.

An n -variate Gaussian normal distribution $\mathcal{N}(\ell, q)$ on \mathbb{R}^n is given by a pair (ℓ, q) where $\ell \in \mathbb{R}[X]_1$ is a linear form and $q \in \mathbb{R}[X]_2$ a positive semi-definite (psd) quadratic form. Some authors require q to be positive definite, but even if q has a nontrivial kernel, the pair (ℓ, q) still defines a normal distribution on the affine subspace given by the *mean vector* $(\ell(e_1), \dots, \ell(e_n))$ (e_i is the i -th coordinate vector) plus the orthogonal complement of the kernel of q (in the maximal degenerate case, i.e. when $q = 0$, this definition gives the Dirac distribution at the mean vector).

The moments of a random variable Y can be used to construct a formal power series

$$\mathbb{E}[\exp(\langle X, Y \rangle)] := \sum_{d=0}^{\infty} \frac{1}{d!} \mathbb{E}[\langle X, Y \rangle^d] \in \mathbb{R}[[X]]$$

which is called the *moment generating series* of Y . The expectation \mathbb{E} is taken X -coefficient-wise over the randomness of Y . For the case of a normally distributed random variable $Y \sim \mathcal{N}(\ell, q)$, this power series takes the very simple and convenient representation

$$(2.1) \quad \mathbb{E}[\exp(\langle X, Y \rangle)] = \exp(\ell + q) = \sum_{d=0}^{\infty} \frac{1}{d!} (\ell + q)^d$$

from which we can read all moments by comparing coefficients. In the case of a centered Gaussian distribution, we have $\ell = 0$ and therefore for each $d \in \mathbb{N}_{>0}$ the moments of degree $2d$ are just simply the coefficients of q^d . All odd-order moments are zero for a centered Gaussian.

A *mixture* Y of m Gaussian random vectors Y_1, \dots, Y_m on \mathbb{R}^n with mixing weights $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{\geq 0}$ satisfying $\sum_{i=1}^m \lambda_i = 1$ is a random vector sampled as follows: From a box containing indices $\{1, \dots, m\}$, draw the index i with probability λ_i and then take a sample of Y_i . It is easy to see that for any integrable function f taking values in \mathbb{R}^n , $\mathbb{E}[f(Y)] = \sum_{i=1}^m \lambda_i \mathbb{E}[f(Y_i)]$. We therefore suggestively denote $Y = \lambda_1 Y_1 \oplus \dots \oplus \lambda_m Y_m$, to remind that on the level of moments and expectations, a mixture random variable is essentially just a convex combination. " \oplus " should not be confused with the actual addition on \mathbb{R}^n . Let the Gaussian random vectors be given by pairs of linear and psd quadratic forms $(\ell_1, q_1), \dots, (\ell_m, q_m)$. The *moment problem for mixtures of Gaussians* asks to obtain these parameter forms given the moments of the mixture up to a certain degree. In the special case of *centered* Gaussians, i.e. $\ell_1 = \dots = \ell_m = 0$, the degree $2d$ moments of $\lambda_1 Y_1 \oplus \dots \oplus \lambda_m Y_m$ attain the form

$$(2.2) \quad \sum_{i=1}^m \lambda_i q_i^d$$

for each $d \in \mathbb{N}_0$. Since $\lambda q^d = (\sqrt[d]{\lambda} q)^d$ for each $\lambda \in \mathbb{R}_{\geq 0}, q \in S^2(\mathbb{C}^n)$, the expression in Equation (2.2) is overparameterized as long as we are only considering moments of one, fixed degree. Let us set the problem in the uniform case, i.e. assume that $\lambda_1 = \dots = \lambda_m = \frac{1}{m}$. Then, the problem corresponds to identifiability for d -th powers of *real psd* quadratic forms.

Remark 2.1. Note that the psd quadratic forms are a (Zariski) dense subset of $S^2(\mathbb{R}^n)$. Since the map $(q_1, \dots, q_m) \mapsto \sum_{i=1}^m q_i^d$ is given by rational polynomials, its image when restricted to real points (or even rational points) is (Zariski) dense in its complex image. Therefore, it suffices to show generic identifiability for d -th powers of complex quadratic forms.

Note that 6 (i.e., $d = 3$) is the smallest degree for which we can hope for identifiability. Indeed, the case of sums of squares ($d = 2$) is not identifiable due to the identity

$$q_1^2 + q_2^2 = \frac{1}{2}(q_1 + q_2)^2 + \frac{1}{2}(q_1 - q_2)^2.$$

Before proving our main result Theorem 1.1 on general identifiability of sextics of certain subgeneric ranks as sums of cubes, we show that indeed it allows us to prove our Corollary 1.2 on identifiability of mixtures of centered Gaussians. We give the reduction for arbitrary d rather than $d = 3$ and remark that it indeed proves 1.2

Theorem 2.2. Let $n, m, d \in \mathbb{N}$ such that generic identifiability holds for d -th powers of quadratic forms in n variables of rank m (cfr. e.g. 1.1 and 4.4)

Let $Y_1 \sim \mathcal{N}(0, q_1), \dots, Y_m \sim \mathcal{N}(0, q_m)$ centered normal distributions given by general psd covariance forms $q_1, \dots, q_m \in S^2(\mathbb{R}^n)$. Let Z_1, \dots, Z_m be any other centered Gaussian random vectors on \mathbb{R}^n such that one of the following holds:

- (a) the uniformly weighted Gaussian mixtures $Y = \frac{1}{m}(Y_1 \oplus \dots \oplus Y_m)$ and $Z = \frac{1}{m}(Z_1 \oplus \dots \oplus Z_m)$ agree on the moments of degree $2d$; or
- (b) for general $\lambda_1, \dots, \lambda_m \in \mathbb{R}_{>0}$ and for $\mu_1, \dots, \mu_m \in \mathbb{R}_{\geq 0}$ both summing up to 1, the Gaussian mixtures $Y = \lambda_1 Y_1 \oplus \dots \oplus \lambda_m Y_m$ and $Z = \mu_1 Z_1 \oplus \dots \oplus \mu_m Z_m$ agree on the moments of degree $2d$ and $2d - 2$.

Then $\{Y_1, \dots, Y_m\} = \{Z_1, \dots, Z_m\}$ and $Y = Z$. In case (b), the corresponding mixing weights are equal, too.

Proof of Corollary 1.2. Let $p_1, \dots, p_m \in S^2(\mathbb{R}^n)$ be quadratic psd forms such that

$$Z_1 \sim \mathcal{N}(0, p_1), \dots, Z_m \sim \mathcal{N}(0, p_m)$$

In case (a), we denote $\lambda_i := \mu_i := \frac{1}{m}$ for each $i \in \{1, \dots, m\}$, while in case (b) we fix $\lambda_1, \dots, \lambda_m$ and μ_1, \dots, μ_m accordingly. Knowing that the degree $2d$ moments of $Y = \lambda_1 Y_1 + \dots + \lambda_m Y_m$ and $Z = \mu_1 Z_1 + \dots + \mu_m Z_m$ are equal, by Equation (2.2) we have

$$(2.3) \quad \sum_{i=1}^m (\sqrt[d]{\lambda_i} q_i)^d = \sum_{i=1}^m (\sqrt[d]{\mu_i} p_i)^d$$

where the quadratic forms $\sqrt[d]{\lambda_i} q_i$ are general for each $i \in \{1, \dots, m\}$. By Theorem 1.1 and Remark 2.1, we get

$$\{\sqrt[d]{\lambda_1} q_1, \dots, \sqrt[d]{\lambda_m} q_m\} = \{\sqrt[d]{\mu_1} p_1, \dots, \sqrt[d]{\mu_m} p_m\}$$

Note that for case (a) this is enough to conclude. In case (b), without loss of generality let us assume that

$$\sqrt[d]{\lambda_1} q_1 = \sqrt[d]{\mu_1} p_1, \dots, \sqrt[d]{\lambda_m} q_m = \sqrt[d]{\mu_m} p_m$$

and write $\alpha_1 := \sqrt[d]{\frac{\lambda_1}{\mu_1}}, \dots, \alpha_m := \sqrt[d]{\frac{\lambda_m}{\mu_m}}$. Since the degree $2d - 2$ moments of Y and Z agree, we have

$$\sum_{i=1}^m \lambda_i q_i^{d-1} = \sum_{i=1}^m \mu_i p_i^{d-1}$$

where the $(d - 1)$ -st powers of the quadratic forms $q_1^{d-1}, \dots, q_m^{d-1}$ are linearly independent. Substituting $p_i = \alpha_i q_i$, we obtain

$$\sum_{i=1}^m \lambda_i q_i^{d-1} = \sum_{i=1}^m \mu_i \alpha_i^{d-1} q_i^{d-1}$$

yielding $\mu_i \left(\frac{\lambda_i}{\mu_i}\right)^{\frac{d-1}{d}} = \lambda_i$, i.e. $\mu_i = \lambda_i$ for each $i \in \{1, \dots, m\}$. ■

2.A. Other types of Gaussian Mixture Problems. Various other special cases of Gaussian Mixture decomposition problems have been studied. In the case of equal mixing weights, if all q_1, \dots, q_m are assumed to be zero, then the problem can be translated into identifiability for special symmetric tensors. For example in [11] the authors showed generic identifiability from moments of degree 3 for all subgeneric ranks. If all quadratics are assumed to be equal (but not necessarily zero), then the problem can still be reduced to tensor decomposition. This is clear from a statistical point of view, but can also be seen algebraically: if $q := q_1 = \dots = q_m$, then the third-order moments attain the form

$$\sum_{i=1}^m \ell_i^3 + q \sum_{i=1}^m \ell_i$$

where the point $\sum_{i=1}^m \ell_i = \mathbb{E}[Y_1 + \dots + Y_m]$ is known, since it is the vector of first-order moments of the mixture. Thus one can shift the space such that $\sum_{i=1}^m \ell_i = 0$ and perform classical tensor decomposition on this kind of Gaussian Mixture problem.

The case of centered Gaussians is significantly more complex and a simple reduction to tensor decomposition is out of reach at the moment, even if Garg, Kayal and Saha [15, Section 1.3] argue that Gaussian Mixtures in full generality might only be a slightly more general class of polynomials than sums of powers of quadratics. Nevertheless, the assumption of centeredness makes the moments easier to handle. For a normally distributed $Y \sim \mathcal{N}(\ell, q)$ the degree 6 moment form would otherwise consists of 4 terms. In order to compute them we would have to look at $k \in \{3, 4, 5, 6\}$ in Equation (2.1), obtaining

$$q^3 + \binom{4}{2} q^2 \ell^2 + \binom{5}{1} q \ell^4 + \ell^6$$

as the form whose coefficients are the degree 6 moments of Y .

3. PRELIMINARIES AND GENERAL NOTATION ON SECANT VARIETIES

After reducing the identifiability problem for mixtures of centered Gaussians to the problem of identifiability of sums of powers of quadratic forms, we recall the basic notations about secant varieties and contact loci that are the basic tools of a geometric approach to the question and will be used in the next section to prove Theorem 1.1.

Notation. For $m \in \mathbb{N}_{>0}$ and elements v_1, \dots, v_m of a vector space, let $\langle v_1, \dots, v_m \rangle$ denote the subspace spanned by them. For projective linear subspaces $\mathbb{P}(V_1), \dots, \mathbb{P}(V_m) \subset \mathbb{P}(V)$ we write $\sum_{i=1}^m \mathbb{P}(V_i)$ for the smallest projective linear subspace containing all of them.

Definition 3.1. (Secant variety) Let W be a variety embedded in an affine or projective space and $m \in \mathbb{N}_{>0}$. The m -th secant variety of W is the Zariski-closure of the union of subspaces spanned by m elements of W , i.e.,

$$\sigma_m(W) = \overline{\bigcup_{x_1, \dots, x_m \in W} \langle x_1, \dots, x_m \rangle}.$$

Notation. Recall that the projective space \mathbb{P}^N is the space of equivalence classes of $\mathbb{C}^N \setminus \{0\}$ with respect to the relation which identifies vectors that are one multiple of each other. For any line through the origin $0 \in \mathbb{C}^{N+1}$ we associate a projective point $[x] \in \mathbb{P}^N$ where x is a non-zero point of the line. Given a subvariety W of the projective space \mathbb{P}^N , let \widehat{W} denote the *affine cone* of W , which is the set of all representatives of projective points of W together with the origin, i.e.,

$$\widehat{W} := \{0\} \cup \bigcup_{[x] \in W} \{x\} \subseteq \mathbb{C}^{N+1}.$$

Remark 3.2. *In terms of affine cones, the secant variety has a convenient parameterization: Let W and m be as in Definition 3.1. Then $\widehat{\sigma_m(W)}$ is the closure of the image of*

$$(3.1) \quad \psi_{m,W}: \widehat{W}^m \rightarrow \widehat{\sigma_m(W)}, (x_1, \dots, x_m) \mapsto \sum_{i=1}^m x_i.$$

From the latter parametrization, it is clear that in order to hope for generic identifiability we need a first necessary condition: $\dim \widehat{W}^m = \dim \widehat{\sigma_m(W)}$. The left-hand side is equal to $m \cdot \dim \widehat{W}$ and is clearly an upper bound for the actual dimension of the (cone of) the m -th secant variety. It is called *expected dimension* and, whenever it is not attained we say that the variety is *m -defective*. The computation of dimensions of secant varieties, and in particular the classification of defective ones, is a difficult challenge in classical algebraic geometry. The following is the main tool to approach the problem, due to Terracini [29], which describes the general tangent space of the secant variety $\sigma_m(V)$ in terms of m general tangent spaces of W .

Notation. Given an affine variety W and a point $x \in W$, $T_x W$ denotes its tangent space at x . If W is a projective variety, embedded in \mathbb{P}^N then, abusing notation, we will also write $T_x W$ for the embedded projective tangent subspace.

Lemma 3.3 (Terracini's Lemma [29]). *Let W be a variety and consider for $m \in \mathbb{N}$ the secant $\sigma_m(W)$. For general points $x_1, \dots, x_m \in W$ and general $x \in \langle x_1, \dots, x_m \rangle \subseteq \sigma_m(W)$, we have that*

$$T_x \sigma_m(W) = \sum_{i=1}^m T_{x_i} W$$

Terracini's Lemma gives us a way to determine whether the map (3.1) is generically finite: a first necessary condition for generic identifiability.

Proposition 3.4. *Let W be an irreducible variety and $m \in \mathbb{N}_0$ such that for general $x_1, \dots, x_m \in W$, the tangent spaces at x_1, \dots, x_m are skew, i.e.*

$$(3.2) \quad \sum_{i=1}^m T_{x_i} \widehat{W} = \bigoplus_{i=1}^m T_{x_i} \widehat{W}$$

Then the map (3.1) is generically finite.

Proof. Let $x_1, \dots, x_m \in W$ be general points. By Lemma 3.3 and generality of x_1, \dots, x_m , the left hand side of Equation (3.2) has the dimension of $\widehat{\sigma_m(W)}$, while $\bigoplus_{i=1}^m T_{x_i} \widehat{W} \cong T_{(x_1, \dots, x_m)} \widehat{W}^m$ has the dimension of \widehat{W}^m . Let $x \in \langle x_1, \dots, x_m \rangle$ be a general point. The fiber dimension formula yields together with Terracini's Lemma (Lemma 3.3) and the assumption,

$$\dim \psi_{m,W}^{-1}(x) = \dim \widehat{W}^m - \dim \widehat{\sigma_m(W)} = \dim \bigoplus_{i=1}^m T_{x_i} \widehat{W} - \dim \sum_{i=1}^m T_{x_i} \widehat{W} = 0.$$

■

Proposition 3.4 gives a tool to examine whether a mixture decomposition problem has only finitely many solutions, basically only requiring us to calculate the dimension of certain vector spaces. Answering the question of identifiability requires further analysis, since the question is a priori not just about the dimension of the generic fibers of $\psi_{m,W}$, but about their cardinality. In other words, we need to show that the map $\psi_{m,W}$ is actually birational. If so, we say that both $\psi_{m,W}$ and $\sigma_m(W)$ are *generically identifiable*.

However, in [6], it is shown that under certain numerical assumptions, proving non m -defectivity implies generic $(m-1)$ -identifiability. We will employ this fact in Section 4.A.

On the other hand, in a series of papers [9, 10, 11], the geometry of the so-called *tangential contact locus* has been used to study identifiability for tensor decompositions. We will follow the same idea in Section 4.B.

Definition 3.5. Let W a variety, $m \in \mathbb{N}_{>0}$ and $x = (x_1, \dots, x_m) \in W^m$ an m -tuple of smooth points of W with skew tangent spaces. The (m -th) *tangential contact locus* $\mathcal{C}_W(x)$ of W at x is the projective subvariety of

$$\Gamma_W(x) := \left\{ y \in W \mid T_y W \subseteq \sum_{i=1}^m T_{x_i} W \right\}.$$

constituting of points $z \in \Gamma_W(x)$ such that the irreducible components of $\Gamma_W(x)$ passing through z contain at least one of x_1, \dots, x_m .

An easy semicontinuity argument enables us to check general identifiability only by studying the tangential contact locus of a specific decomposition $t = x_1 + \dots + x_m$. Indeed a more general statements holds, see for example [10, Proposition 2.3]

4. IDENTIFIABILITY FOR POWERS OF FORMS

Let $V_{k,d} = \{q^d \mid q \in \mathbb{P}(S^k(\mathbb{C}^n))\}$ denote the projective variety of d -th powers of degree- k forms. We suppress the dependency on $n \in \mathbb{N}$.

Proposition 4.1. *For $k, d \in \mathbb{N}$, the map*

$$\iota : \mathbb{P}(S^k(\mathbb{C}^n)) \rightarrow V_{k,d} \subseteq \mathbb{P}(S^{kd}(\mathbb{C}^n)), p \mapsto p^d$$

is an embedding.

Proof. The map ι can be regarded as the following composition:

$$\mathbb{P}(S^k(\mathbb{C}^n)) \xrightarrow{\nu_d} \mathbb{P}(S^d(S^k(\mathbb{C}^n))) \xrightarrow{\pi_E} \mathbb{P}(S^{kd}(\mathbb{C}^n))$$

where:

- ν_d is the d -th Veronese embedding sending linear forms to their degree- d power;
- π_E is the orthogonal linear projection induced by the decomposition

$$S^d(S^k(\mathbb{C}^n)) = S^{kd}(\mathbb{C}^n) \oplus E$$

where E is the degree- d part of the ideal of the k -th Veronese embedding of $\mathbb{P}(\mathbb{C}^n)$. The center of the projection E does not intersect the 2-nd secant variety $\sigma_2(\nu_d(\mathbb{P}(S^k(\mathbb{C}^n))))$. Indeed, all forms in the latter secant variety is either of the form $\ell^{d-1}m$ or $\ell^d + m^d$, where ℓ, m are linear forms, and they are both completely reducible, since $\ell^d + m^d = \prod_{i=1}^d (\ell + \zeta^i m)$ where $\zeta = e^{\frac{2\pi i}{d}}$ is a primitive d -th root of unity. However, the Veronese variety is irreducible and is not contained in any hyperplane. Therefore, $\sigma_2(\nu_d(\mathbb{P}(S^k(\mathbb{C}^n)))) \cap E = \emptyset$ and the projection π_E restricted to the image $\nu_d(\mathbb{P}(S^k(\mathbb{C}^n)))$ is an isomorphism and the composition $\pi_E \circ \nu_d$ is an embedding. ■

We prove general identifiability for the m -th secant variety of $V_{k,d}$ with two approaches.

4.A. From non-defectivity to identifiability. In [6], Casarotti and Mella derive general identifiability as a consequence of the next-order secant having expected dimension.

Theorem 4.2. [6, Introduction] *Let W be a smooth variety of dimension $n \in \mathbb{N}$ and let $m \in \mathbb{N}$. Assume that the m -th secant variety is of (expected) dimension $m(n+1) - 1$ and $m > 2n$. Then W is $(m-1)$ -identifiable.*

By simple computation of differentials, it is immediate to notice that the tangent space to $V_{k,d}$ at $q = p^d$ is given by $\{hp^{d-1} \mid h \in S^k(\mathbb{C}^n)\}$. Therefore, by Terracini's Lemma (Lemma 3.3), in order to prove that secant varieties of $V_{k,d}$ have the expected dimension, we only have to show that the tangent spaces $T_{p_1^d} V_{k,d}, \dots, T_{p_m^d} V_{k,d}$ are skew for a general choice of the p_i 's. That is equivalent to say that the degree- kd part of the ideal $(p_1^{d-1}, \dots, p_m^{d-1})$ has maximal dimension. This is implied by a more general fact related to Fröberg's Conjecture on Hilbert series of general ideals.

Given a homogeneous ideal $I \subset \mathbb{C}[X]$, the Hilbert series of the associated quotient ring is

$$\text{HS}(\mathbb{C}[X]/I; T) = \sum_{i \in \mathbb{N}} \dim(\mathbb{C}[X]_i/I_i) T^i \in \mathbb{N}[[T]]$$

where $I_i := I \cap \mathbb{C}[X]_i$. Fröberg's Conjecture [12] says that given a general ideal $I = (g_1, \dots, g_m) \subset \mathbb{C}[X]$ with $\deg(g_i) = d_i$, the Hilbert series is given by the formula

$$(4.1) \quad \left[\frac{\prod_{i=1}^m (1 - T^{d_i})}{(1 - T)^n} \right]$$

where $[\cdot]$ means that the power series obtained by the fraction is truncated before the first non-positive coefficient. In [25, Conjecture 2], it is conjectured that, whenever $\deg(p_i) > 1$, the ideal $(p_1^{d-1}, \dots, p_m^{d-1})$ has the Hilbert series (4.1) for $d_i = (d-1)k$ for a general choice of the p_i 's, namely

$$\left[(1 - T^{(d-1)k})^m \cdot \sum_{j \in \mathbb{N}} \binom{n-1+j}{n-1} T^j \right].$$

If the latter holds, then it is immediate to see that the coefficient of T^{dk} is

$$\binom{n-1+kd}{n-1} - m \binom{n-1+k}{n-1} = \dim \mathbb{C}[X]_{kd} - m \cdot \dim \mathbb{C}[X]_k,$$

i.e., the tangent spaces $T_{p_1^d} V_{k,d}, \dots, T_{p_m^d} V_{k,d}$ are skew.

For fixed positive integers a, h , Nenashév showed in [24, Theorem 1] that the coefficient of T^{a+h} of the Hilbert series of an ideal I is as prescribed by (4.1) whenever $I = (g_1, \dots, g_m)$ with $\deg(g_i) = a$, where the g_i 's are chosen generically from a nonempty variety $\mathcal{D} \subset S^d(\mathbb{C}^n)$ that is closed under linear transformation, and $m \leq \frac{\dim \mathbb{C}[X]_{a+h}}{\dim \mathbb{C}[X]_h} - \dim \mathbb{C}[X]_h$.

In conclusion, by applying the latter result for $a = (d-1)k, h = k$ and \mathcal{D} is the tangential variety of $V_{k,d}$, i.e., $\mathcal{D} = \{p^{d-1}h \mid p, h \in S^k\}$, we immediately deduce the following.

Theorem 4.3. *The dimension of the m -th secant variety of $V_{k,d}$ is as expected, i.e.,*

$$\dim \sigma_m V_{k,d} = m \cdot \dim \mathbb{C}[X]_k - 1$$

for $m \leq \frac{\dim \mathbb{C}[X]_{kd}}{\dim \mathbb{C}[X]_k} - \dim \mathbb{C}[X]_k$.

Therefore, by Theorem 4.2 and since generic m -identifiability implies generic $(m-1)$ -th identifiability, we have the following identifiability result.

Corollary 4.4. *The m -th secant variety of $V_{k,d}$ is generically m -identifiable for $m \leq \frac{\dim \mathbb{C}[X]_{kd}}{\dim \mathbb{C}[X]_k} - \dim \mathbb{C}[X]_k - 1$ provided that $2(\dim \mathbb{C}[X]_k - 1) < \frac{\dim \mathbb{C}[X]_{kd}}{\dim \mathbb{C}[X]_k} - \dim \mathbb{C}[X]_k$.*

In particular, in the $(k, d) = (2, 3)$ case, we obtain the condition (1) of the main Theorem 1.1. Indeed, note that the condition required by Corollary 4.4, i.e.,

$$2 \left(\binom{n+1}{2} - 1 \right) < \frac{\binom{n+5}{6}}{\binom{n+1}{2}} - \binom{n+1}{2},$$

holds if and only if $n > 16$.

Corollary 4.5 (Theorem 1.1, Condition (1)). *Let $n, m \in \mathbb{N}$ such that $n > 16$. Then, for $m \leq \binom{n+5}{6} / \binom{n+1}{2} - \binom{n+1}{2} - 1$ and general $q_1, \dots, q_m \in S^2(\mathbb{C}^n)$ and general $t \in \langle q_1^3, \dots, q_m^3 \rangle$ there is a unique representation of t as sum m cubes of quadratic forms.*

Corollary 4.4 provides results also for sums of higher powers of high degree forms. In the case of power of quadratics ($k = 2$), the condition required by Corollary 4.4 reduces to

$$3 \binom{n+1}{2}^2 - 2 \binom{n+1}{2} < \binom{n-1+2d}{2d}$$

and Theorem 1.3 follows. The latter inequality holds for pairs (n, d) in the region represented in Figure 2.

4.B. Tangential contact locus for cubes of quadratics. The second approach relies on an earlier result due to Chiantini and Ottaviani [9], which provides generic identifiability as a consequence of a dimension argument for the tangential contact locus.

A semicontinuity argument enables us to check general identifiability only by studying the tangential contact locus of a specific decomposition $t = x_1 + \dots + x_m$. Indeed, the statements below relate both via another question of dimension.

Proposition 4.6 ([10, Proposition 2.3]). *Let W be an irreducible, nondegenerate variety of dimension $n \geq 2$, which is not m -defective. If the generic element of $\sigma_m(W)$ is not identifiable, then for general $x \in W^m$ and each $i \in \{1, \dots, m\}$, the tangential contact locus to W at x must contain a curve through x_i .*

Theorem 4.7 ([9, Proposition 2.4]). *Let W be a nondegenerate, irreducible smooth variety and $m \in \mathbb{N}_{>0}$. Consider the following statements:*

- (i) *The m -th secant map $\psi_{m,W}$ is generically identifiable.*
- (ii) *For every m general points $x_1, \dots, x_m \in W$, $T_{x_1}W, \dots, T_{x_m}W$ are skew spaces and the dimension of $\mathcal{C}_W(x_1, \dots, x_m)$ at every x_i is zero.*
- (iii) *There exist m specific points $x_1, \dots, x_m \in W$ with skew tangent spaces*

$$T_{x_1}W, \dots, T_{x_m}W$$

such that the dimension of $\mathcal{C}_W(x_1, \dots, x_m)$ at a specific x_i is zero.

Then we have (iii) \implies (ii) \implies (i).

In order to prove the identifiability result, we will construct a specific set of quadratic forms q_1, \dots, q_m , where $m = \binom{n}{2} + 1$, such that the tangential contact locus at q_1, \dots, q_m consists only of the points q_1, \dots, q_m with skew tangent spaces. This proves by Theorem 4.7 that the secant of rank up to $\binom{n}{2} + 1$ is identifiable, for *any* $n \in \mathbb{N}$. We will use variables $X = (X_1, \dots, X_n)$ as local affine coordinates. Given $q_1, \dots, q_m \in \mathbb{C}[X]_2$ we denote, by abuse of notation, with $\widehat{\mathcal{C}}(q_1, \dots, q_m)$ the affine cone of the preimage of the tangential contact locus at the points $[q_1^3], \dots, [q_m^3]$ via the map ι from proposition 4.1. Similarly, we denote by $\widehat{\Gamma}(q_1, \dots, q_m) := \iota^{-1}(\Gamma_{V_{2,3}}(\widehat{[q_1^3]}, \dots, \widehat{[q_m^3]}))$ the preimage via ι of $\Gamma_{V_{2,3}}$ at q_1, \dots, q_m , cf. definition 3.5. This notation suppresses the dependency on the variety $V := V_{2,3}$, which depends on the number n of variables. Making the expression for tangent spaces at V explicit, we have:

$$\widehat{\Gamma}(q_1, \dots, q_m) = \left\{ p \in \mathbb{C}[X]_2 \mid \forall h \in \mathbb{C}[X]_2 : \exists h_1, \dots, h_m \in \mathbb{C}[X]_2 : p^2 h = \sum_{i=1}^m q_i^2 h_i \right\}.$$

Definition 4.8. For $i, j \in \{1, \dots, n\}$ with $i < j$, define

$$(4.2) \quad q_{ij} := (X_i + X_j)^2$$

and let

$$\mathcal{B}_n := \{(X_i + X_j)^2 \mid i, j \in \{1, \dots, n\}, i < j\} \cup \{X_1^2\}$$

We call \mathcal{B}_n the *binomial set* of quadratics in dimension n . Up to relabeling we can write $\mathcal{B}_n = \{q_1, \dots, q_{\binom{n}{2}+1}\}$, where the order of the elements is arbitrary.

Remark 4.9. *It is an easy consequence of the previous proposition that the following equality holds for $n \geq 2$:*

$$\mathcal{B}_{n-1} \cup \{4X_1^2\} = \{p(X_1, \dots, X_{n-1}, X_1) \mid p \in \mathcal{B}_n\}$$

Theorem 4.10. *The tangent spaces at elements of the binomial set \mathcal{B}_n are skew, i.e.*

$$T_{q_1} \widehat{V} + \dots + T_{q_{\binom{n}{2}+1}} \widehat{V} = T_{q_1} \widehat{V} \oplus \dots \oplus T_{q_{\binom{n}{2}+1}} \widehat{V}$$

Proof. We proceed by induction on n . For $n \leq 5$, we verify the statement on a computer. The code may be found on GitHub [28]. Therefore we can assume that $n \geq 6$ and that the claim on \mathcal{B}_k is true for all $k < n$. Let $h_1, h_{ij} \in S^2(\mathbb{C}^n)$, where $i, j \in \{1, \dots, n\}$ and $i < j$. Suppose that

$$(4.3) \quad 0 = h_1 X_1^4 + \sum_{1 \leq i < j \leq n} h_{ij} (X_i + X_j)^4$$

We have to show that $h_1 = h_{ij} = 0$ for all $1 \leq i < j \leq n$. Denote $h_{ji} := h_{ij}$. Since \mathcal{B}_n is symmetric under permutations of $\{X_2, \dots, X_n\}$, we can interchange any two variables not equal to X_1 . Thus without loss of generality the only two cases to consider are $(i, j) = (2, 3)$ and $(i, j) = (1, 2)$. Let us first consider the case $(i, j) = (2, 3)$. Since $n \geq 6$, we may apply the substitution

$$\varphi_4: \mathbb{C}[X] \rightarrow \mathbb{C}[X_1, \dots, X_3, X_5, \dots, X_n], X_4 \mapsto X_1$$

to reduce to a case with one variable less. We obtain

$$(4.4) \quad 0 = \varphi_4(h_1) X_1^4 + \sum_{\substack{1 \leq k < l \leq n \\ 4 \notin \{k, l\}}} \varphi_4(h_{kl}) (X_k + X_l)^4 + \sum_{k=1}^n \varphi_4(h_{4k}) (X_k + X_1)^4$$

Now note that the form $(X_2 + X_3)^4$ can only occur on the first summation, yielding $\varphi_4(h_{23}) = 0$. Therefore by construction $(X_1 - X_4)$ divides the quadratic form h_{23} . Repeating this same argument with the substitutions

$$\begin{aligned} \varphi_5: X_5 &\mapsto X_1 \\ \varphi_6: X_6 &\mapsto X_1 \end{aligned}$$

yields that $(X_1 - X_5)$ and $(X_1 - X_6)$ divide h_{23} , too. Since these linear forms are coprime, $(X_1 - X_4)(X_1 - X_5)(X_1 - X_6)$ must divide h_{23} , which for degree reasons is only possible if $h_{23} = 0$. By symmetry of \mathcal{B}_n , we get that $h_{ij} = 0$ for all pairs $\{i, j\}$ not containing 1. Thus Equation (4.3) simplifies to

$$(4.5) \quad 0 = h_1 X_1^4 + \sum_{j=1}^n h_{1j} (X_1 + X_j)^4$$

As for the $(i, j) = (1, 2)$ case: If h_{12} were not the zero form, then $h_{12}(X_1 + X_2)^4$ would contain a monomial of degree at least 4 in X_2 . Since all other addends in Equation (4.5) can only contain monomials of degree at most 2 in X_2 , the terms of degree at least 4 in X_2 from $h_{12}(X_1 + X_2)^4$ could not cancel with any other addend from (4.5). After a short argument left to the reader, this forces $h_{12} = 0$ and by symmetry thus $h_{13} = \dots = h_{1n} = 0$. Finally, we also must have $h_1 = 0$, as it is the only remaining term in (4.5). \blacksquare

Now we show that the tangential contact locus for the binomial set is zero dimensional at each point of the binomial set.

Theorem 4.11. *For $n \in \mathbb{N}$, and each $q \in \{q_1, \dots, q_{\binom{n}{2}+1}\}$, locally around q , $\widehat{\mathcal{C}}(q_1, \dots, q_{\binom{n}{2}+1})$ only contains points from the line $\mathbb{C}q$.*

Proof. We use affine notation and proceed by induction on the number n of variables. The base cases $n \leq 5$ were verified on a computer, see [28].

Thus let us assume $n \geq 6$. As \mathcal{B}_n is invariant under permutations of X_2, \dots, X_n , it suffices to show the claim at $q \in \{X_1^2, (X_1 + X_2)^2, (X_2 + X_3)^2\}$. In particular, we may assume that q is a polynomial in X_1, X_2, X_3 . As we work locally around q , it does not matter whether we show the statement for $\widehat{\Gamma}$ or $\widehat{\mathcal{C}}$, cf. definition 3.5. We thus have to show that there exists a neighbourhood $\mathcal{U} \subseteq \mathbb{C}[X]_2$ of q such that $\mathcal{U} \cap \widehat{\Gamma}(q_1, \dots, q_m) \subseteq \mathbb{C}q$. Consider the substitution

$$\varphi: \mathbb{C}[X] \rightarrow \mathbb{C}[X_1, \dots, X_{n-1}]$$

that maps X_n to X_1 . Note $\varphi(q) = q$, as $n \geq 6$. By induction hypothesis, we know there exists a neighbourhood $\mathcal{V} \subseteq \mathbb{C}[X_1, \dots, X_{n-1}]_2$ of q such that $\mathcal{V} \cap \widehat{\Gamma}(\mathcal{B}_{n-1}) \subseteq \mathbb{C}q$. This means that for all

$$p \in \varphi^{-1}(\mathcal{V}) \cap \widehat{\Gamma}(q_1, \dots, q_m)$$

there is $\lambda \in \mathbb{C}$ such that $\varphi(p) = \lambda q$. In other words,

$$(X_1 - X_n)|(p - \lambda q)$$

Repeating the same argument with the substitution φ' that maps X_{n-1} to X_1 , we obtain another neighbourhood \mathcal{V}' with the property that for each

$$p \in \varphi'^{-1}(\mathcal{V}') \cap \widehat{\Gamma}(q_1, \dots, q_m)$$

it holds that $\varphi'(p) = \lambda' q$.

Let $\mathcal{U} = \varphi^{-1}(\mathcal{V}) \cap \varphi'^{-1}(\mathcal{V}')$, then for each $p \in \mathcal{U} \cap \widehat{\Gamma}(q_1, \dots, q_m)$, we can find $\lambda, \lambda' \in \mathbb{C}$ and linear forms $\ell, \ell' \in \mathbb{C}[X]_1$ such that

$$(4.6) \quad \lambda q + \ell(X_1 - X_n) = p = \lambda' q + \ell'(X_1 - X_{n-1})$$

Finally, we have that ℓ has to be a polynomial in the variables $\{X_1, X_{n-1}\}$: Indeed, if a variable X_j for some $j \notin \{1, n-1\}$ occurred in ℓ , then the monomial $X_j X_n$ on the left hand side of (4.6) could not cancel with any other terms on the left-hand side, but does also not occur on the right hand side. It follows that p is a polynomial in $\{X_1, X_2, X_3, X_{n-1}, X_n\}$. Thus we reduced to the case of 5 variables and proved the claim. \blacksquare

- Remark 4.12.** (a) *Our results do not imply that the mixture of cubes of quadratics $\sum_{q \in \mathcal{B}_n} q^3$ has a unique decomposition as a sum of $\binom{n}{2} + 1$ cubes of quadratics! In [11], the authors consider some sufficient criteria for the identifiability of specific tensors that maybe, albeit with unnegligible effort, could be transferred to the setting of cubes of quadratics. We did not do any work regarding specific identifiability for cubes of quadratics.*
- (b) *We verify Theorem 4.10 and Theorem 4.11 for $n = 5$ on a computer. The code is publicly available on GitHub, see [28]. This base case can be verified using only methods of Numerical Linear Algebra (such as determining dimensions of certain vector spaces of polynomials) and should therefore be easy to reproduce independently.*

REFERENCES

- [1] Carlos Amendola, Jean-Charles Faugere, and Bernd Sturmfels. Moment varieties of gaussian mixtures. *Journal of Algebraic Statistics*, 7(1), 2016.
- [2] Anima Anandkumar, Rong Ge, Daniel Hsu, Sham M. Kakade, and Matus Telgarsky. Tensor decompositions for learning latent variable models. *Journal of Machine Learning Research*, 2012.
- [3] Alessandra Bernardi and Daniele Taufer. Waring, tangential and cactus decompositions. *Journal de Mathématiques Pures et Appliquées*, 143:1–30, Nov 2020.
- [4] Jeff Bezanson, Alan Edelman, Stefan Karpinski, and Viral B Shah. Julia: A fresh approach to numerical computing. *SIAM Review*, 59(1):65–98, 2017.
- [5] Maria Chiara Brambilla and Giorgio Ottaviani. On the alexander–hirschowitz theorem. *Journal of Pure and Applied Algebra*, 212(5):1229–1251, 2008.
- [6] Alex Casarotti and Massimiliano Mella. From non-defectivity to identifiability. *Journal of the European Mathematical Society*, 2022.
- [7] Luca Chiantini and Ciro Ciliberto. Weakly defective varieties. *Transactions of the American Mathematical Society*, 354(1):151–178, 2002.
- [8] Luca Chiantini and Ciro Ciliberto. On the dimension of secant varieties. *Journal of the European Mathematical Society*, 12(5):1267–1291, 2010.
- [9] Luca Chiantini and Giorgio Ottaviani. On generic identifiability of 3-tensors of small rank. *SIAM Journal on Matrix Analysis and Applications*, 33(3):1018–1037, 2012.
- [10] Luca Chiantini, Giorgio Ottaviani, and Nick Vannieuwenhoven. An algorithm for generic and low-rank specific identifiability of complex tensors. *SIAM Journal on Matrix Analysis and Applications*, 35(4):1265–1287, 2014.
- [11] Luca Chiantini, Giorgio Ottaviani, and Nick Vannieuwenhoven. On generic identifiability of symmetric tensors of subgeneric rank. *Transactions of the American Mathematical Society*, 369(6):4021–4042, 2016.

- [12] Ralf Fröberg. An inequality for hilbert series of graded algebras. *Mathematica Scandinavica*, 56(2):117–144, 1985.
- [13] Ralf Fröberg, Giorgio Ottaviani, and Boris Shapiro. On the waring problem for polynomial rings. *Proceedings of the National Academy of Sciences*, 109(15):5600–5602, 2012.
- [14] Francesco Galuppi and Massimiliano Mella. Identifiability of homogeneous polynomials and cremona transformations. *Journal für die reine und angewandte Mathematik (Crelles Journal)*, 2019(757):279–308, 2019.
- [15] Ankit Garg, Neeraj Kayal, and Chandan Saha. 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.
- [16] Rong Ge, Qingqing Huang, and Sham M. Kakade. Learning mixtures of gaussians in high dimensions. *Proceedings of the forty-seventh annual ACM symposium on Theory of Computing*, 2015.
- [17] Rong Ge and Tengyu Ma. Decomposing Overcomplete 3rd Order Tensors using Sum-of-Squares Algorithms. In Naveen Garg, Klaus Jansen, Anup Rao, and José D. P. Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2015)*, volume 40 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 829–849, Dagstuhl, Germany, 2015. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik.
- [18] Richard Harshman. Foundations of the parafac procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics*, 16, 1970.
- [19] A Hirschowitz, J Alexander, and A Hirschowitz. Polynomial interpolation in several variables. *Journal of Algebraic Geometry*, 4(4):201–222, 1995.
- [20] Joe Kileel, Matthew Trager, and Joan Bruna. On the expressive power of deep polynomial neural networks. *Advances in neural information processing systems*, 32, 2019.
- [21] Benoît Legat, Sascha Timme, and Robin Deits. Juliaalgebra/multivariatepolynomials.jl: v0.3.18, July 2021.
- [22] Samuel Lundqvist, Alessandro Oneto, Bruce Reznick, and Boris Shapiro. On generic and maximal k-ranks of binary forms. *Journal of Pure and Applied Algebra*, 223(5):2062–2079, 2019.
- [23] Tengyu Ma, Jonathan Shi, and David Steurer. Polynomial-time tensor decompositions with sum-of-squares. In Irit Dinur, editor, *FOCS*, pages 438–446. IEEE Computer Society, 2016.
- [24] Gleb Nenashev. A note on fröberg’s conjecture for forms of equal degrees. *Comptes Rendus Mathématique*, 355(3):272–276, 2017.
- [25] Lisa Nicklasson. On the hilbert series of ideals generated by generic forms. *Communications in Algebra*, 45(8):3390–3395, 2017.
- [26] Karl Pearson. 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:pp. 1–47+405, 1900.
- [27] Bruce Reznick. Equal sums of two cubes of quadratic forms. *International Journal of Number Theory*, 17(03):761–786, 2021.
- [28] Alexander Taveira Blomenhofer. Base case computation for: Sums of third Powers of Quadratics are generically Identifiable up to quadratic Rank, April 2022. Available at <https://github.com/a441/cubes-of-quadratics>.
- [29] A. Terracini. Sulle V_k che rappresentano più di $\frac{k(k-1)}{2}$ equazioni di laplace linearmente indipendenti. *Rend. Circ. Mat. Palermo*, 33:176–186, 1912.

UNIVERSITY OF KONSTANZ, GERMANY, FACHBEREICH MATHEMATIK UND STATISTIK, D-78457 KONSTANZ, GERMANY

Email address: alexander.taveira-blomenhofer@uni-konstanz.de

UNIVERSITÀ DI TRENTO, VIA SOMMARIVE, 14 - 38123 POVO (TRENTO), ITALY

Email address: alex.casarotti@unitn.it

UNIVERSITY OF KONSTANZ, GERMANY, FACHBEREICH MATHEMATIK UND STATISTIK, FACH D 197 D-78457 KONSTANZ, GERMANY

Email address: mateusz.michalek@uni-konstanz.de

UNIVERSITÀ DI TRENTO, VIA SOMMARIVE, 14 - 38123 POVO (TRENTO), ITALY

Email address: alessandro.oneto@unitn.it