Moment Identifiability of Homoscedastic Gaussian Mixtures

Daniele Agostini[1] · Carlos Améndola[2] · Kristian Ranestad[3]

Abstract

We consider the problem of identifying a mixture of Gaussian distributions with the same unknown covariance matrix by their sequence of moments up to certain order. Our approach rests on studying the moment varieties obtained by taking special secants to the Gaussian moment varieties, defined by their natural polynomial parametrization in terms of the model parameters. When the order of the moments is at most three, we prove an analogue of the Alexander–Hirschowitz theorem classifying all cases of homoscedastic Gaussian mixtures that produce defective moment varieties. As a consequence, identifiability is determined when the number of mixed distributions is smaller than the dimension of the space. In the two-component setting, we provide a closed form solution for parameter recovery based on moments up to order four, while in the one-dimensional case we interpret the rank estimation problem in terms of secant varieties of rational normal curves.

Communicated by Peter Bürgisser.

✉ Carlos Améndola
  carlos.amendola@tum.de

  Daniele Agostini
  daniele.agostini@math.hu-berlin.de

  Kristian Ranestad
  ranestad@math.uio.no

[1] Humboldt Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany

[2] Technische Universität München, Boltzmannstraße 3, 85748 Garching, Germany

[3] Universitetet i Oslo, Postboks 1053 Blindern, 0316 Oslo, Norway

Published online: 06 July 2020

🖄 Springer

1 Introduction

In the context of algebraic statistics [19], moments of probability distributions have recently been explored from an algebraic and geometric point of view [1,4,11,13]. The key point for this connection is that in many cases the sets of moments define algebraic varieties, hence called *moment varieties*. In the case of moments of mixture distributions, there is a natural correspondence to secant varieties of the moment varieties. Studying geometric invariants such as their dimension reveals properties such as model identifiability. One of the main applications for statistical inference is in the context of the method of moments, which matches the distribution's moments to moment estimates obtained from a sample.

Gaussian mixtures are a prominent statistical model with multiple applications (see [3] and references therein). They are probability distributions on $\mathbb{R}^n$ with a density that is a convex combination of Gaussian densities:

$$\lambda_1 f_{\mathcal{N}(\mu_1, \Sigma_1)}(x) + \cdots + \lambda_k f_{\mathcal{N}(\mu_k, \Sigma_k)}(x) \tag{1}$$

where $\mu_1, \ldots, \mu_k \in \mathbb{R}^n$ are the k means, $\Sigma_1, \ldots, \Sigma_k \in \mathrm{Sym}^2(\mathbb{R}^n)$ are the covariance matrices, and the $0 \leq \lambda_i \leq 1$ with $\lambda_1 + \cdots + \lambda_k = 1$ are the mixture weights.

The starting point is thus the Gaussian moment variety $\mathcal{G}_{n,d}$, as introduced in [4], whose points are the vectors of all moments of order at most d of an n-dimensional Gaussian distribution. The moments corresponding to the mixture density (1) form the secant variety $\mathrm{Sec}_k(\mathcal{G}_{n,d})$, and identifiability in this general setting was the focus of [5].

In this work, we study special families of Gaussian mixtures, called homoscedastic mixtures, where all the Gaussian components share the same covariance matrix. In other words, a *homoscedastic* Gaussian mixture has a density of the form

$$\sum_{i=1}^{k} \lambda_i f_{\mathcal{N}(\mu_i, \Sigma)}(x) \tag{2}$$

where the Gaussian probability densities $f_{\mathcal{N}(\mu_i, \Sigma)}(x)$ have all different means μ_i and *same* covariance matrix Σ. The moments, up to order d, of homoscedastic Gaussian mixtures are still polynomials in the parameters (the means and the covariance matrix), and form the moment variety $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$. This is a set of special k-secants inside the secant variety $\mathrm{Sec}_k(\mathcal{G}_{n,d})$.

The main question we are concerned with is: when can a general homoscedastic k-mixture of n-dimensional Gaussians be identified by its moments of order d? More precisely, denote by $\Theta_{n,k}^H$ the parameter space of means, covariances and mixture weights for homoscedastic mixtures, and the moment map by

$$M_{n,k,d} : \Theta_{n,k}^H \to \mathrm{Sec}_k^H(\mathcal{G}_{n,d}). \tag{3}$$

The mixture parameters of a point on the moment variety $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$ can be uniquely recovered if the fiber of the moment map (3) is a singleton up to natural permutations

of the parameters. If this happens for a general point on the moment variety, we say that the mixture is *rationally identifiable* from its moments up to order d. If the fiber of a general point is finite, we say that we have *algebraic identifiability*. The parameters are *not identifiable* if the general fiber of the moment map has positive dimension.

If the dimension of the parameter space is larger than the dimension of the space of moments, then one may expect any moment to lie on the moment variety. Clearly, the fiber of the moment map must have positive dimension and we cannot have identifiability. We therefore distinguish the unexpected cases: when the dimension of the moment variety is less than the dimension of both the parameter space and the moment space, then we say that the moment variety $\operatorname{Sec}_k^H(\mathcal{G}_{n,d})$ is *defective*. In particular, defectivity implies non-identifiability.

We illustrate with an example:

Example 1 Let $n = 2, k = 2$ and $d = 3$. That is, we consider moments up to order three for the homoscedastic mixture of two Gaussians in $\mathbb{R}^2$. The Gaussian moment variety $\mathcal{G}_{2,3}$ is 5-dimensional with 2 parameters for the mean vector and 3 for the symmetric covariance matrix. The parameters for the homoscedastic mixture are two mean vectors $\mu_1 = \begin{pmatrix} \mu_{11} \\ \mu_{12} \end{pmatrix}$ and $\mu_2 = \begin{pmatrix} \mu_{21} \\ \mu_{22} \end{pmatrix}$, the common covariance $\Sigma = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}$ and the mixture weight λ of the first component, in total $2 \times 2 + 3 + 1 = 8$ parameters. On the other hand, there are 9 bivariate moments up to order 3. Explicitly, the map is:

$$m_{10} = \lambda\mu_{11} + (1 - \lambda)\mu_{21}$$
$$m_{01} = \lambda\mu_{12} + (1 - \lambda)\mu_{22}$$
$$m_{20} = \lambda(\mu_{11}^2 + \sigma_{11}) + (1 - \lambda)(\mu_{21}^2 + \sigma_{11})$$
$$m_{02} = \lambda(\mu_{12}^2 + \sigma_{22}) + (1 - \lambda)(\mu_{22}^2 + \sigma_{22})$$
$$m_{11} = \lambda(\mu_{11}\mu_{12} + \sigma_{12}) + (1 - \lambda)(\mu_{21}\mu_{22} + \sigma_{12})$$
$$m_{30} = \lambda(\mu_{11}^3 + 3\sigma_{11}\mu_{11}) + (1 - \lambda)(\mu_{21}^3 + 3\sigma_{11}\mu_{21})$$
$$m_{03} = \lambda(\mu_{12}^3 + 3\sigma_{22}\mu_{12}) + (1 - \lambda)(\mu_{22}^3 + 3\sigma_{22}\mu_{22})$$
$$m_{21} = \lambda(\mu_{11}^2\mu_{12} + \sigma_{11}\mu_{12} + 2\sigma_{12}\mu_{11}) + (1 - \lambda)(\mu_{21}^2\mu_{22} + \sigma_{11}\mu_{22} + 2\sigma_{12}\mu_{21})$$
$$m_{12} = \lambda(\mu_{11}\mu_{12}^2 + \sigma_{22}\mu_{11} + 2\sigma_{12}\mu_{12}) + (1 - \lambda)(\mu_{21}\mu_{22}^2 + \sigma_{22}\mu_{21} + 2\sigma_{12}\mu_{22})$$

Since there are more moments than parameters, one would expect that the mixture parameters can be recovered. However, the dimension of $\operatorname{Sec}_2^H(\mathcal{G}_{2,3})$ equals 7. This is one less than the expected dimension of 8. Therefore, it is defective and there is no algebraic identifiability. This means that the method of moments is doomed to fail in this setting. However, if one measures moments up to order $d = 4$, it is possible to uniquely recover the mixture parameters.

As is often observed [1,4,13], a change of coordinates to cumulants tends to yield simpler representations and faster computations. This is the case here and hence we also study the *cumulant varieties* of the homoscedastic Gaussian mixtures. For Example 1, the moment variety in cumulant coordinates is simply the cone over a twisted cubic curve (see Example 5). This is not a coincidence, as is shown in Sect. 3.

Our main results, Theorems 2 and 3, identify the defective homoscedastic moment varieties when $d = 3$ and show that the homoscedastic moment variety is not defective

when $k \leq n + 1$. These are analogues of the Alexander–Hirschowitz theorem on secant-defective Veronese varieties [2].

This paper is organized as follows: In Sect. 2 we present the connection between moments and cumulants. The moment varieties corresponding to homoscedastic mixtures are defined in Sect. 3. In Sect. 4 we give general algebraic identifiability considerations and do a careful analysis of the subcases $d = 3$, $k = 2$ and $n = 1$. Finally, we conclude with a summary of results and list further research directions.

2 Moments and Cumulants

To get started, we make some remarks about moments and cumulants from an algebraic perspective. To a sufficiently integrable random variable X on $\mathbb{R}^n$, associate its moments $m_{a_1,\ldots,a_n}[X]$ and cumulants $\kappa_{a_1,\ldots,a_n}[X]$ through the generating functions in $\mathbb{R}[\![u_1, \ldots, u_n]\!]$:

$$
\begin{aligned}
M_X(u) &= \sum_{(a_1,\ldots,a_n)} m_{a_1,\ldots,a_n}[X] \frac{u_1^{a_1} \ldots u_n^{a_n}}{a_1! \ldots a_n!} \\
K_X(u) &= \sum_{(a_1,\ldots,a_n)} \kappa_{a_1,\ldots,a_n}[X] \frac{u_1^{a_1} \ldots u_n^{a_n}}{a_1! \ldots a_n!}.
\end{aligned}
\tag{4}
$$

The information obtained from moments is equivalent to that from cumulants, since they are obtained from one another through the simple transformations

$$
M_X(u) = \exp(K_X(u)), \qquad K_X(u) = \log(M_X(u))
\tag{5}
$$

which are well-defined, because the 0-th moment is always one, whereas the 0-th cumulant is always zero: $m_0[X] = 1$, $\kappa_0[X] = 0$ for every random variable X. In particular, moments and cumulants take values in the affine hyperplanes $\mathbb{A}_n^M$ and $\mathbb{A}_n^K$ of $\mathbb{R}[\![u_1, \ldots, u_n]\!]$ defined by

$$
\mathbb{A}_n^M = \{m_0 = 1\}, \qquad \mathbb{A}_n^K = \{\kappa_0 = 0\}.
\tag{6}
$$

We call these hyperplanes the *moment space* and the *cumulant space*.

Taking only moments up to order d, replace the ring $\mathbb{R}[\![u_1, \ldots, u_n]\!]$ of power series with the truncated ring $\mathbb{R}[\![u_1, \ldots, u_n]\!]/(u_1, \ldots, u_n)^{d+1}$, and everything goes through. In particular, there is an analogous definition of the affine hyperplanes $\mathbb{A}_{n,d}^M$ and $\mathbb{A}_{n,d}^K$ which we also call moment space and cumulant space.

Example 2 (**Dirac distribution**) Let $\mu = (\mu_1, \ldots, \mu_n)$ in $\mathbb{R}^n$ be a point. The Dirac distribution δ_μ with center μ on $\mathbb{R}^n$ is given by

$$
\int_{\mathbb{R}^n} f(x)\delta_\mu(x) := f(\mu).
\tag{7}
$$

If X is a random variable on $\mathbb{R}^n$ with this distribution, its moment-generating function is

$$M_X(u) = \mathbb{E}[e^{u^t X}] = e^{u^t \mu} = \sum_{(a_1,\ldots,a_n)} \mu_1^{a_1} \cdots \mu_n^{a_n} \frac{u_1^{a_1} \cdots u_n^{a_n}}{a_1! \ldots a_n!}. \tag{8}$$

The moments of X are monomials evaluated at μ. On the other hand, for the cumulant generating function

$$K_X(u) = \log M_X(u) = \log e^{u^t \mu} = u^t \mu = \mu_1 u_1 + \cdots + \mu_n u_n, \tag{9}$$

the linear cumulants coincide with the coordinates of μ, and the higher order cumulants are all zero.

This has an immediate translation into algebro-geometric terms: the parameter space for all Dirac distributions is the space $\mathbb{R}^n$, and the image of the moment map of degree d, $M : \mathbb{R}^n \to \mathbb{A}_{n,d}^M$ is the affine d-th Veronese variety $V_{n,d} \subseteq \mathbb{A}_{n,d}^M$. On the other hand, the image of the cumulant map $K : \mathbb{R}^n \to \mathbb{A}_{n,d}^K$ is the linear subspace given by $\{\kappa_2 = \kappa_3 = \cdots = \kappa_d = 0\}$, where κ_i is the degree i-part of an element in $\mathbb{A}_{n,d}^K$.

Example 3 (**Gaussian distribution**) Let $\mu \in \mathbb{R}^n$ be a point, and $\Sigma \in \mathrm{Sym}^2 \mathbb{R}^n$ an $n \times n$ symmetric and positive-definite matrix. The Gaussian distribution on $\mathbb{R}^n$ with mean μ and covariance matrix Σ is given by the density

$$f_{(\mu,\Sigma)}(x) := \frac{1}{\sqrt{\det(2\pi\Sigma)}} e^{-\frac{1}{2}(x-\mu)^t \Sigma^{-1}(x-\mu)}. \tag{10}$$

If $X \sim \mathcal{N}(\mu, \Sigma)$ is a Gaussian random variable with these parameters, its moment-generating function and cumulant-generating function are given by

$$M_X(u) = e^{u^t \mu + \frac{1}{2} u^t \Sigma u}, \qquad K_X(u) = u^t \mu + \frac{1}{2} u^t \Sigma u. \tag{11}$$

The Gaussian moment variety $\mathcal{G}_{n,d} \subseteq \mathbb{A}_{n,d}^M$ consists of all Gaussian moments up to order d. Observe that the corresponding cumulant variety is given simply by the linear subspace $\{\kappa_3 = \cdots = \kappa_d = 0\} \subseteq \mathbb{A}_{n,d}^K$.

While our focus is on Gaussian distributions, our approach applies to general location families that admit moment and cumulant varieties. We illustrate this with the next example.

Example 4 (**Laplace distribution**) The (symmetric) multivariate Laplace distribution has a location parameter $\mu \in \mathbb{R}^n$ and a covariance parameter Σ, a positive-definite $n \times n$ matrix. Its density function involves the modified Bessel function of the second kind (see [12, Chapter 5]), but it can be defined via its simpler moment generating function:

$$M_X(u) = \frac{\exp(u^t \mu)}{1 - \frac{1}{2} u^t \Sigma u}, \qquad K_X(u) = u^t \mu - \log\left(1 - \frac{1}{2} u^t \Sigma u\right) \tag{12}$$

with radius of convergence such that $|u^t \Sigma u| < 2$.

Moments and cumulants up to order $d = 3$ match with the Gaussian case. Also note that when $\Sigma = 0$, the Dirac moment generating function is recovered. However, when $d \geq 4$, the Laplace cumulants are no longer a linear space in the cumulant space.

The multiplicative structure of the power series ring $\mathbb{R}[[u_1, \cdots, u_n]]$ makes it particularly suitable to independence statements with respect to moments. Indeed, if X, Y are two independent random variables on $\mathbb{R}^n$, then

$$M_{X+Y}(u) = \mathbb{E}[e^{u^t(X+Y)}] = \mathbb{E}[e^{u^t X} e^{u^t Y}] = \mathbb{E}[u^t X] \cdot \mathbb{E}[u^t Y] = M_X(u) \cdot M_Y(u).$$

With cumulants it is even simpler: it holds that

$$K_{X+Y}(u) = \log(M_{X+Y}) = \log(M_X) + \log(M_Y) = K_X(u) + K_Y(u).$$

The group of affine transformations $\mathrm{Aff}(\mathbb{R}^n)$ acts naturally on both moments and cumulants: indeed, for any $A \in GL(n, \mathbb{R})$ and $b \in \mathbb{R}^n$ and a random variable X on $\mathbb{R}^n$,

$$M_{AX+b}(u) = M_{AX}(u) \cdot M_b(u) = \mathbb{E}[e^{u^t AX}] \cdot \mathbb{E}[e^{u^t b}] = e^{u^t b} \cdot M_X(A^t u)$$

and

$$K_{AX+b}(u) = \log(M_{AX+b}(u)) = \log(e^{u^t b} M_X(A^t u)) = u^t b + K_X(A^t u).$$

In particular, note that translations correspond simply to translations in cumulant coordinates, whereas they induce a more complicated expression in moment coordinates.

3 Homoscedastic Secants

When Karl Pearson introduced Gaussian mixtures to model subpopulations of crabs [18], he also proposed the *method of moments* in order to estimate the parameters. The basic idea is to compute sample moments from observed data, and match them to the distribution's moments expressed in terms of the unknown parameters. The method of moments estimates are the parameters that solve these equations. This is a classical estimation method in statistics; a good survey is [16], and a recent 'denoised' version for Gaussian mixtures is [21].

The method of moments is very friendly for mixture models because computing moments of mixture densities is straightforward, since for every measurable function $g \colon \mathbb{R}^n \to \mathbb{R}$

$$\int_{\mathbb{R}^n} g(x) \left(\sum_{i=1}^k \lambda_i f_{(\mu_i, \Sigma_i)}(x) \right) \mathrm{d}x = \sum_{i=1}^k \lambda_i \int_{\mathbb{R}^n} g(x) f_{(\mu_i, \Sigma_i)}(x) \mathrm{d}x, \qquad (13)$$

and thus, the moments are just linear combinations of the corresponding Gaussian moments.

As hinted in the introduction, this discussion can be rephrased in geometric terms: let $\mathcal{G}_{n,d} \subseteq \mathbb{A}_{n,d}^M$ be the Gaussian moment variety on $\mathbb{R}^n$ of order d. Then, the moments of mixtures of Gaussians are linear combinations of points in $\mathcal{G}_{n,d}$, so that their corresponding variety is the k-th secant variety $\mathrm{Sec}_k(\mathcal{G}_{n,d})$.

The densities of *homoscedastic* Gaussian mixtures, where the Gaussian components share a common covariance matrix, have the form:

$$\lambda_1 f_{(\mu_1, \Sigma)}(x) + \cdots + \lambda_k f_{(\mu_k, \Sigma)}(x) \tag{14}$$

where the $\mu_i \in \mathbb{R}^n$ are the mean parameters, the $\Sigma \in \mathrm{Sym}^2 \mathbb{R}^n$ is the common covariance parameters, and the $\lambda_i \in \mathbb{R}$ with $\lambda_1 + \cdots + \lambda_k = 1$ are the mixture parameters. Thus, the parameter space for homoscedastic mixtures is

$$\begin{aligned}
\Theta_{n,k}^H &:= (\mathbb{R}^n)^{\times k} \times \mathbb{R}^{k-1} \times \mathrm{Sym}^2 \mathbb{R}^n \\
&= \{((\mu_1, \ldots, \mu_k), (\lambda_1, \ldots, \lambda_k), \Sigma) \mid \lambda_1 + \cdots + \lambda_k = 1\},
\end{aligned} \tag{15}$$

and it has dimension

$$\dim \Theta_{n,k}^H = nk + k - 1 + \frac{n(n+1)}{2} = (n+1)\left(k + \frac{n}{2}\right) - 1. \tag{16}$$

The moment map for homoscedastic mixtures is then an algebraic map

$$M_{n,k,d} : \Theta_{n,k}^H \to \mathbb{A}_{n,d}^M.$$

Points on the image, the moments of homoscedastic mixtures, are linear combinations of points in $\mathcal{G}_{n,d} \subseteq \mathbb{A}_{n,d}^M$ which share the same covariance matrix.

Definition 1 The *homoscedastic k-secant variety*, denoted $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$, is the image of the moment map $M_{n,k,d}$. The *fiber dimension* $\Delta_{n,k,d}^H$ is the general fiber dimension of the map $M_{n,k,d}$,

$$\Delta_{n,k,d}^H = \dim \Theta_{n,k}^H - \dim \mathrm{Sec}_k^H(\mathcal{G}_{n,d}). \tag{17}$$

We say that $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$ is *algebraically identifiable* if $\Delta_{n,k,d}^H = 0$.

The feasibility of the method of moments is based on computing points on the fibers of the moment map $M_{n,k,d}$. Algebraic identifiability of $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$ means that a general homoscedastic Gaussian mixture in the homoscedastic k-secant variety is identifiable from its moments up to order d in the sense that only finitely many Gaussian mixture distributions share the same moments up to order d, whereas we reserve the term *rationally identifiable* if a general fiber consists of a single point, up to label swapping. In case the general fiber is not finite, then it is positive-dimensional, there is no identifiability of the parameters from the moments up to order d, and a higher order is needed for identifiability (cf. Remark 4 and [4, Problem 17]).

Since the dimension of $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$ is always bounded by the dimension of the ambient space $\mathbb{A}_{n,d}^M$, there is a simple estimate for the fiber dimension:

Lemma 1 *For all n, d, k it holds that*

$$\Delta_{n,k,d}^H \geq \max\left\{(n+1)\left(k + \frac{n}{2}\right) - \binom{n+d}{d}, 0\right\}. \qquad (18)$$

Proof The moment space $\mathbb{A}_{n,d}^M$ is an affine hyperplane inside the vector space $\mathbb{R}[[u_1, \ldots, u_n]]/(u_1, \ldots, u_n)^{d+1}$; hence, it has dimension

$$\dim \mathbb{A}_{n,d}^M = \dim \mathbb{R}[[u_1, \ldots, u_n]]/(u_1, \ldots, u_n)^{d+1} - 1 = \binom{n+d}{d} - 1. \quad (19)$$

Since $\mathrm{Sec}_k^H(\mathcal{G}_{n,d}) \subseteq \mathbb{A}_{n,d}^M$, note that

$$\Delta_{n,k,d}^H = \dim \Theta_{n,k}^H - \dim \mathrm{Sec}^H(\mathcal{G}_{n,d}) \geq \dim \Theta_{n,h}^H - \dim \mathbb{A}_{n,d}^M \qquad (20)$$

which is exactly the inequality in the statement. $\qquad\square$

We expect that in general situations the inequality (18) is in fact an equality. Hence, define the *defect* to be

$$\delta_{n,k,d}^H := \Delta_{n,k,d}^H - \max\left\{(n+1)\left(k + \frac{n}{2}\right) - \binom{n+d}{d}, 0\right\}. \qquad (21)$$

We say that $\mathrm{Sec}_k^H(\mathcal{G}_{n,d})$ is *defective* if $\delta_{n,k,d}^H > 0$. As observed earlier, defectivity implies non-identifiability.

3.1 Cumulant Representation

Let us explore how homoscedastic secants become simpler in cumulant coordinates, and how this representation can be used to check identifiability.

First, rephrase the situation in terms of random variables: let $Z = Z_\Sigma$ be a Gaussian random variable with mean 0 and covariance matrix Σ, and let $B = B_{(\mu_1, \ldots, \mu_k), (\lambda_1, \ldots, \lambda_k)}$ an independent random variable with distribution given by a mixture of Dirac distributions:

$$\lambda_1 \delta_{\mu_1}(x) + \cdots + \lambda_k \delta_{\mu_k}(x). \qquad (22)$$

Then, the random variable $Z + B$ has density given by the homoscedastic mixture (1). Moreover, if $m = \mu_1 \lambda_1 + \cdots + \mu_k \lambda_k$ is the mean of B, we write $B = A + m$, where A is a centered mixture of Dirac distributions.

One can compute cumulants of this random variable as follows:

$$K_{B+Z}(u) = K_B(u) + K_Z(u) = K_A(u) + m^t u + \frac{1}{2} u^t \Sigma u \qquad (23)$$

and this suggests to parametrize the homoscedastic secants in cumulant coordinates as follows:

$$K : \Theta_{n,k}^0 \times \mathbb{R}^n \times \operatorname{Sym}^2 \mathbb{R}^n \to \mathbb{A}_{n,d}^K, \qquad (A, m, \Sigma) \mapsto K_A(u) + m^t u + \frac{1}{2} u^t \Sigma u \tag{24}$$

where $\Theta_{n,k}^0$ parametrizes the centered mixtures of Dirac distributions

$$\Theta_{n,k}^0 = \{(\mu_1, \ldots, \mu_k), (\lambda_1, \ldots, \lambda_k) \mid \lambda_1 \mu_1 + \cdots + \lambda_k \mu_k = 0, \; \lambda_1 + \cdots + \lambda_k = 1\}$$

The *cumulant homoscedastic secant variety* $\log(\operatorname{Sec}_k^H(\mathcal{G}_{n,d}))$ is the image of the map K. Since in this variety, one can freely translate by the elements in $\mathbb{R}^n$ and $\operatorname{Sym}^2(\mathbb{R}^n)$, the first cumulants and the second cumulants can take any value. The constraints are in the cumulants of order three and higher. We summarize this discussion in the following lemma.

Lemma 2 *Let $\mathbb{A}_{n,d}^{K,3}$ be the space of cumulants of order at least three and at most d, let*

$$\phi_{n,k,d} : \Theta_{n,k}^0 \to \mathbb{A}_{n,d}^{K,3}, \qquad A \mapsto K_A(u)_3 + K_A(u)_4 + \cdots + K_A(u)_d \tag{25}$$

be the cumulant map and let $C_{n,k,d}^0$ denote the closure $\overline{\phi_{n,k,d}(\Theta_{n,k}^0)}$. Then, the cumulant homoscedastic secant variety $\log(\operatorname{Sec}_k^H(\mathcal{G}_{n,d}))$ is a cone over $C_{n,k,d}^0$.

Remark 1 In particular, the equations for the cumulant homoscedastic secant variety $\log(\operatorname{Sec}^H(\mathcal{G}_{n,d}))$ inside $\mathbb{A}_{n,d}^K$ are exactly the same as the equations for $C_{n,k,d}^0$ inside $\mathbb{A}_{n,d}^{K,3}$.

The fiber dimension $\Delta_{n,k,d}^H$ can also be computed as the fiber dimension of the map $\phi_{n,k,d}$:

Lemma 3 *The fiber dimension $\Delta_{n,k,d}^H$ is equal to the fiber dimension of $\phi_{n,k,d}$. In other words*

$$\Delta_{n,k,d}^H = \dim \Theta_{n,k}^0 - \dim C_{n,k,d}^0 = (k-1)(n+1) - \dim C_{n,k,d}^0. \tag{26}$$

Proof The fiber dimension is the difference $\dim \Theta_{n,k}^H - \dim \log(\operatorname{Sec}_k^H(\mathcal{G}_{n,d}))$. We know that $\Theta_{n,k}^H \cong \Theta_{n,k}^0 \times \mathbb{R}^n \times \operatorname{Sym}^2 \mathbb{R}^n$. Moreover, Lemma 2 says that $\log(\operatorname{Sec}_k^H(\mathcal{G}_{n,d}))$ is the cone over $C_{n,k,d}^0$, which is precisely $\mathbb{R}^n \times \operatorname{Sym}^2 \mathbb{R}^n \times C_{n,k,d}^0$, so that the first equality follows. For the second equality, the dimension of $\Theta_{n,k}^0$ can be computed as $nk + k - 1 - n = (n+1)(k-1)$. □

Example 5 ($n = k = 2$, $d = 3$) Revisiting Example 1 from the introduction, we concluded that $\operatorname{Sec}_2^H(\mathcal{G}_{2,3}) \subset \mathbb{A}_{2,3}^M \cong \mathbb{A}^9$ is expected to be a hypersurface but it is actually

of codimension 2. The ideal of $\mathrm{Sec}_2^H(\mathcal{G}_{2,3})$ is Cohen–Macaulay and determinantal (generated by the maximal minors of a 6×5-matrix) as described in [4, Proposition 19]. The homoscedastic cumulant variety $\log(\mathrm{Sec}_2^H(\mathcal{G}_{2,3}))$ is defined by the vanishing of the 2×2 minors of

$$\begin{pmatrix} k_{30} & k_{21} & k_{12} \\ k_{21} & k_{12} & k_{03} \end{pmatrix}.$$

Note that indeed the first- and second-order cumulants $k_{10}, k_{01}, k_{20}, k_{11}, k_{22}$ do not appear in the equations above, so that the cumulant variety is the cone over the twisted cubic curve.

Remark 2 To estimate the mixture parameters from the cumulants, it is enough to consider the map $\phi_{n,k,d}$ of Lemma 2. Indeed, suppose that we have a homoscedastic mixture with parameters $(((\lambda_1, \ldots, \lambda_k), (\mu_1, \ldots, \mu_k)), m, \Sigma) \in \Theta_{n,k}^0 \times \mathbb{R}^n \times \mathrm{Sym}^2 \mathbb{R}^n$ and suppose that its cumulants are known, so that in polynomial form

$$\kappa_1(u) = m^t u$$
$$\kappa_2(u) = K_A(u)_2 + \frac{1}{2} u^t \Sigma u$$
$$\kappa_3(u) = K_A(u)_3 \qquad\qquad . \tag{27}$$
$$\kappa_4(u) = K_A(u)_4$$
$$\vdots$$

Then, to recover the parameters one can first try to recover the λ_i and the μ_i from the cumulants of order three and higher, and then compute m and Σ from the cumulants of order one and two.

3.2 Veronese Secants

We briefly observe that we can recast the above discussion in a way that makes apparent the connection to mixtures of Dirac distributions and, hence, to secants of Veronese varieties. To work with classical secant varieties, this time we work in moment coordinates. Now, every homoscedastic mixture is the distribution of a random variable of the form $Z + B$, where B is a mixture of Dirac distributions and Z is a centered Gaussian of covariance Σ, independent from B. Thus, the moment generating function of this variable is

$$M_{Z+B}(u) = M_Z(u) M_B(u) = e^{\frac{1}{2} u^t \Sigma u} \cdot M_B(u). \tag{28}$$

Therefore, the role of the covariance parameter is decoupled from the others: In particular, for $\Sigma = 0$, one obtains the moment variety for mixtures of Dirac distributions. When restricting to moments $M(u)_d$ of degree at most d, this is precisely the k-secants to the Veronese variety $\mathrm{Sec}_k(\mathcal{V}_{n,d})$. The additive group $\mathrm{Sym}^2 \mathbb{R}^n$ acts on the moment

space $\mathbb{A}_{n,d}^M$ by

$$\text{Sym}^2 \mathbb{R}^n \times \mathbb{A}_{n,d}^M \to \mathbb{A}_{n,d}^M, \qquad (\Sigma, M(u)_d) \mapsto e^{\frac{1}{2}u^t \Sigma u} \cdot M(u)_d \qquad (29)$$

and so (28) says that $\text{Sec}_k^H(\mathcal{G}_{n,d})$ is the union of all the orbits of the points in $\text{Sec}_k(\mathcal{V}_{n,d})$ under this action.

This is useful because we can exploit well-known results on secants of Veronese varieties to address identifiability. First, let $\Delta_{n,k,d}^{\mathcal{V}}$ denote the fiber dimension of the k-secants to the Veronese variety $\text{Sec}_k(\mathcal{V}_{n,d}) \subseteq \mathbb{A}_{n,d}^M$: by definition, this is

$$\Delta_{n,k,d}^{\mathcal{V}} := nk + k - 1 - \dim \text{Sec}_k(\mathcal{V}_{n,d}). \qquad (30)$$

A basic estimate for the dimension of $\text{Sec}_k(\mathcal{V}_{n,d})$ is given by the dimension of the ambient space $\dim \mathbb{A}_{n,d}^M = \binom{n+d}{d} - 1$, hence

$$\Delta_{n,k,d}^{\mathcal{V}} \geq \max \left\{ (n+1)k - \binom{n+d}{d}, 0 \right\} \qquad (31)$$

so that we can define the defect of the k-secants to the Veronese variety as

$$\delta_{n,k,d}^{\mathcal{V}} := \Delta_{n,k,d}^{\mathcal{V}} - \max \left\{ (n+1)k - \binom{n+d}{d}, 0 \right\}. \qquad (32)$$

This number was famously computed by Alexander and Hirschowitz [2], see also [7]:

Theorem 1 (Alexander–Hirschowitz) *The defect for the Veronese variety is always zero, except in the following exceptional cases*

$$\begin{aligned}
&d = 2, 2 \leq k \leq n & &\Delta_{n,k,2}^{\mathcal{V}} = \frac{k(k-1)}{2} \\
&n = 2, d = 4, k = 5 & &\delta_{2,5,4}^{\mathcal{V}} = 1 \\
&n = 3, d = 4, k = 9 & &\delta_{3,9,4}^{\mathcal{V}} = 1 \\
&n = 4, d = 3, k = 7 & &\delta_{4,7,3}^{\mathcal{V}} = 1 \\
&n = 4, d = 4, k = 14 & &\delta_{4,14,4}^{\mathcal{V}} = 1
\end{aligned} \qquad (33)$$

Moreover, for a general point $M(u) \in \text{Sec}_k(\mathcal{V}_{n,d})$, consider the closed subset of $\text{Sym}^2 \mathbb{R}^n$ given by

$$D(M) := \{\Sigma \in \text{Sym}^2 \mathbb{R}^n \mid e^{\frac{1}{2}u^t \Sigma u} \cdot M(u) \in \text{Sec}_k(\mathcal{V}_{n,d})\}. \qquad (34)$$

We have the following relation between the fiber dimensions (17) and (30):

Proposition 1 *It holds that*

$$\Delta^H_{n,k,d} = \Delta^{\mathcal{V}}_{n,k,d} + \dim D(M) \tag{35}$$

where $M \in \mathrm{Sec}_k(\mathcal{V}_{n,d})$ is a general point.

Proof By the previous discussion, the moment map for homoscedastic mixtures factors as a composition of two surjective maps

$$\Theta^H_{n,k} \to \mathrm{Sym}^2(\mathbb{R}^n) \times \mathrm{Sec}_k(\mathcal{V}_{n,d}) \to \mathrm{Sec}^H_k(\mathcal{G}_{n,d}). \tag{36}$$

Hence, the fiber dimension of the composite map is the sum of the fiber dimensions of the two factors. For the first one this is $\Delta^{\mathcal{V}}_{n,k,d}$, so it remains to consider the second. Denote the second factor by $\rho\colon \mathrm{Sym}^2(\mathbb{R}^n) \times \mathrm{Sec}_k(\mathcal{V}_{n,d}) \to \mathrm{Sec}^H_k(\mathcal{V}_{n,d})$ and let $(\Sigma_o, M_o(u)) \in \mathrm{Sym}^2(\mathbb{R}^n) \times \mathrm{Sec}_k(\mathcal{V}_{n,d})$ be a general point. The fiber is

$$\begin{aligned}
\rho^{-1}(\rho(\Sigma_o, M_o(u))) &= \left\{ (\Sigma, M(u)) \,|\, e^{\frac{1}{2}u^t \Sigma u} \cdot M(u) = e^{\frac{1}{2}u^t \Sigma_o u} \cdot M_o(u) \right\} \\
&= \{ (\Sigma, M(u)) \,|\, M(u) = e^{\frac{1}{2}u^t(\Sigma_o - \Sigma)u} \cdot M_o(u) \} \\
&\cong \{ \Sigma \in \mathrm{Sym}^2(\mathbb{R}^n) \,|\, e^{\frac{1}{2}u^t(\Sigma_o - \Sigma)u} \cdot M_o(u) \in \mathrm{Sec}_k(\mathcal{V}_{n,d}) \} \\
&= \Sigma_o - \{ \Sigma' \in \mathrm{Sym}^2(\mathbb{R}^n) \,|\, e^{\frac{1}{2}u^t \Sigma' u} \cdot M_o(u) \in \mathrm{Sec}_k(\mathcal{V}_{n,d}) \} \\
&\cong D^K(M_o),
\end{aligned}$$

concluding the proof. $\qquad\square$

Remark 3 In the range $(n+1)\left(k + \frac{n}{2}\right) \le \binom{n+d}{d}$ where we expect identifiability for homoscedastic Gaussian mixtures, we see that $\Delta^H_{n,k,d} = \delta^H_{n,k,d}$, and Alexander–Hirschowitz says that $\Delta^{\mathcal{V}}_{n,k,d} = \delta^{\mathcal{V}}_{n,k,d} = 0$. Hence, Proposition 1 yields

$$\delta^H_{n,k,d} = \dim D(M) \tag{37}$$

4 Moment Identifiability

Now we start to determine identifiability in various cases. To do so, it is convenient to change notation slightly. Up to now, we have identified moments and cumulants with their corresponding generating functions. In the next sections, it is useful to identify the parameters with polynomials as well. We replace the location parameter $\mu = (\mu_1, \dots, \mu_n)$ with the corresponding linear polynomial $u^t \mu = \mu_1 u_1 + \cdots + \mu_n u_n$ and we replace the covariance parameter Σ with the quadric $\frac{1}{2}u^t \Sigma u$. Of course, the two representations are equivalent, but the polynomial formalism is better suited to the cumulant space and the moment space. In particular, the linear polynomials live in the dual vector space $V = \mathrm{Hom}(\mathbb{R}^n, \mathbb{R})$, whereas the quadratic polynomials live in $\mathrm{Sym}^2 V$.

The next inequality reflects the fact that increasing the order of moments (or cumulants) measured results in better identifiability:

Lemma 4 *The fiber dimensions of general fibers of $M_{n,k,d}$ and $M_{n,k,d+1}$ satisfy:*

$$\Delta_{n,k,d}^{H} \geq \Delta_{n,k,d+1}^{H}. \tag{38}$$

Proof By definition, the fiber dimension $\Delta_{n,k,d}^{H}$ is the dimension of a general nonempty fiber of the moment map $M_{n,k,d}: \Theta_{n,k,d}^{H} \to \mathbb{A}_{n,d}^{M}$. However, this map is the composition of the map $M_{n,k,d+1}: \Theta_{n,k,d}^{H} \to \mathbb{A}_{n,d+1}^{M}$ and the projection map $\mathbb{A}_{n,d+1}^{M} \to \mathbb{A}_{n,d}^{M}$, that forgets the moments of order $d+1$, so the conclusion follows. $\square$

Remark 4 Since Gaussian mixtures are identifiable from finitely many moments (see, e.g., [4]), the sequence

$$\Delta_{n,k,1}^{H} \geq \Delta_{n,k,2}^{H} \geq \cdots \geq \Delta_{n,k,d}^{H} \geq \Delta_{n,k,d+1}^{H} \geq \cdots$$

must stabilize at 0 for some large enough d.

The following observation is less trivial. It allows a reduction to the case $n = k - 1$.

Proposition 2 *Suppose that $d \geq 3$ and $n \geq k - 1$, then*

$$\Delta_{n,k,d}^{H} = \Delta_{k-1,k,d}^{H}. \tag{39}$$

Proof Use Lemma 3, which says that the fiber dimension $\Delta_{n,k,d}^{H}$ is equal to the fiber dimension of the map

$$\phi_{n,k,d}: \Theta_{n,k}^{0} \to \mathbb{A}_{n,d}^{K,3}. \tag{40}$$

This dimension can be computed by looking at the differential of the map at a general point. The parameter space is defined as

$$\Theta_{n,k}^{0} = \left\{ ((\lambda_1, \ldots, \lambda_k), (L_1, \ldots, L_k)) \in \mathbb{R}^k \times V^k \mid \sum_{i=1}^{k} \lambda_i = 1, \ \sum_{i=1}^{k} \lambda_i L_i = 0 \right\}.$$

Let $p = ((\lambda_1, \ldots, \lambda_k), (L_1, \ldots, L_k)) \in \Theta_{n,k}^{0}$ be a general point. Then, the tangent space to $\Theta_{n,k}^{0}$ at the point is given by

$$T_p \Theta_{n,k}^{0} = \left\{ ((\varepsilon_i)_{i=1}^{k}, (H_i)_{i=1}^{k}) \in \mathbb{R}^k \times V^k \mid \sum_{i=1}^{k} \varepsilon_i = 0, \ \sum_{i=1}^{k} (\varepsilon_i L_i + \lambda_i H_i) = 0 \right\}.$$

The fiber dimension of $\phi_{n,k,d}$ coincides with the dimension of the kernel of the differential $d\phi_{n,k,d}$ at the general point p. In particular, since the point is general and $n \geq k - 1$, we can suppose that $L_i = u_i$ for $i = 1, \ldots, k - 1$ and that all the

λ_i are nonzero. In particular L_k is a linear combination of $u_1, \ldots, u_{k-1}$. Now, we claim that if $((\varepsilon_1, \ldots, \varepsilon_k), (H_1, \ldots, H_k))$ is in the kernel of $d\phi_{n,k,d}$ then the only variables appearing in the H_i are $u_1, \ldots, u_{k-1}$. If this is true, then we are done, because the kernel of $d\phi_{n,k,d}$ coincides with the kernel of $d\phi_{k-1,k,d}$ at the point $((\lambda_1, \ldots, \lambda_k), (L_1, \ldots, L_k)) \in \Theta^0_{k-1,k}$.

To prove the claim, observe that the map is given by the cumulant functions $\phi_{n,k,d} = (\kappa_3, \kappa_4, \ldots, \kappa_d)$, so the kernel of $d\phi_{n,k,d}$ equals the intersection of the kernels of the $d\kappa_i$ for $i = 3, \ldots, d$. Therefore, it is enough to prove the analogous claim for the kernel of the differential $d\kappa_3$ of κ_3. Since the first moment is zero by construction, the third cumulant coincides with the third moment

$$\kappa_3 = \lambda_1 L_1^3 + \cdots + \lambda_k L_K^3. \tag{41}$$

Hence, the differential is the linear map

$$d\kappa_{3,p} \colon T_p \Theta^0_{n,k} \to \mathbb{A}^{K,3}_{n,d}, \quad ((\varepsilon_1, \ldots, \varepsilon_k), (H_1, \ldots, H_k)) \mapsto \sum_{i=1}^k (3\lambda_i H_i + \varepsilon_i L_i) L_i^2$$

and if $((\varepsilon_1, \ldots, \varepsilon_k), (H_1, \ldots, H_k))$ is in the kernel, then it must be that

$$\sum_{i=1}^k h_i L_i^2 = 0, \quad \text{where } h_i = 3\lambda_i H_i + \varepsilon_i L_i. \tag{42}$$

Since $\lambda_k \neq 0$, this is equivalent to $\sum_{i=1}^k h_i (\lambda_k L_i)^2 = 0$ and since $\lambda_1 L_1 + \cdots + \lambda_k L_k = 0$, we see that

$$\sum_{i=1}^k h_i (\lambda_k L_i)^2 = \sum_{i=1}^{k-1} h_i (\lambda_k L_i)^2 + h_k (\lambda_k L_k)^2 = \sum_{i=1}^{k-1} h_i (\lambda_k L_i)^2 + h_k \left(\sum_{i=1}^{k-1} \lambda_i L_i \right)^2$$

$$= \sum_{i=1}^{k-1} (\lambda_k^2 h_i + \lambda_i^2 h_k) L_i^2 + 2h_k \left(\sum_{1 \leq i < j \leq k-1} \lambda_i \lambda_j L_i L_j \right).$$

By assumption $L_i = u_i$ for $i = 1, \ldots, k-1$, so this last expression is equal to zero if and only if

$$\sum_{i=1}^{k-1} (\lambda_k^2 h_i + \lambda_i^2 h_k) u_i^2 = -2h_k \left(\sum_{1 \leq i < j \leq k-1} \lambda_i \lambda_j u_i u_j \right). \tag{43}$$

If this is true, then h_k uses only the variables $u_1, \ldots, u_{k-1}$. Indeed, if some other variable, say y, appears in h_k, then on the right-hand side there is the monomial $y u_1 u_2$, while there is no such a monomial on the left-hand side. Likewise, if the variable y appears in one of the h_i for $i = 1, \ldots, k-1$: then on the left-hand side there would

be a monomial of the form yu_i^2, while there is no such monomial on the right hand side.

Hence, the h_i are polynomials in the $u_1, \ldots, u_k$, and, by definition of the h_i, it follows that the same holds for the H_i. This proves the claim and the result follows. □

4.1 Moments Up to Order $d = 3$

When $d = 3$ we determine the defect $\delta_{n,k,3}^H$ and the fiber dimension $\Delta_{n,k,3}^H$ of the map

$$\phi_{n,k,3} : \Theta_{n,k}^0 \to \mathbb{A}_{n,3}^{K,3}$$

for each n and k, and use Lemma 3. When $d = 3$, the space $\mathbb{A}_{n,3}^{K,3}$ is identified with the space $\mathrm{Sym}^3 V$ of homogeneous polynomials of degree three, and as noted in the proof of Proposition 2, the third cumulants coincide with the third moments, so that:

$$\phi_{n,k,3} : \Theta_{n,k}^0 \to \mathrm{Sym}^3 V \qquad ((L_1, \ldots, L_k), (\lambda_1, \ldots, \lambda_k)) \mapsto \lambda_1 L_1^3 + \cdots + \lambda_k L_k^3.$$

We compute the closure $C_{n,k,3}^0$ of the image.

Lemma 5 *The set $C_{n,k,3}^0$ is the Zariski closure of*

$$\{H_1(u)^3 + \cdots + H_k(u)^3 \mid H_1(u), \ldots, H_k(u) \in \mathbb{R}^n \text{ linearly dependent }\}. \qquad (44)$$

Proof Recall that

$$\Theta_{n,k}^0 = \{((L_i)_{i=1}^k, (\lambda_i)_{i=1}^k)) \in V^k \times \mathbb{R}^{k-1} | \lambda_1 + \cdots + \lambda_k = 1, \lambda_1 L_1 + \cdots + \lambda_k L_k = 0\}.$$

To compute the Zariski closure, suppose that all the λ_i are strictly positive, so that in particular we can write

$$L_k = -\frac{\lambda_1}{\lambda_k} L_1 - \cdots - \frac{\lambda_{k-1}}{\lambda_k} L_{k-1}. \qquad (45)$$

Since cubic roots are well defined over $\mathbb{R}$,

$$\lambda_1 L_1^3 + \cdots + \lambda_k L_k^3 = \lambda_1 L_1^3 + \cdots + \lambda_{k-1} L_{k-1}^3 - \lambda_k \left(\frac{\lambda_1}{\lambda_k} L_1 + \cdots + \frac{\lambda_{k-1}}{\lambda_k} L_{k-1} \right)^3$$

$$= H_1^3 + \cdots + H_{k-1}^3 + H_k^3$$

where $H_i := \sqrt[3]{\lambda_i} L_i$ for $i = 1, \ldots, k-1$, and $H_k := -\sum_{i=1}^{k-1} \left(\frac{\sqrt[3]{\lambda_i}}{\sqrt[3]{\lambda_k}} \right)^2 H_i$, using the equality $\sqrt[3]{\lambda_k} \frac{\lambda_i}{\lambda_k} = \left(\frac{\sqrt[3]{\lambda_i}}{\sqrt[3]{\lambda_k}} \right)^2 \sqrt[3]{\lambda_i}$. In particular, this shows immediately that $\lambda_1 L_1^3 + \cdots + \lambda_k L_k^3$ can be written as a sum of cubic powers of linearly dependent linear forms.

For the converse, let $H_1, \ldots, H_k$ be linearly dependent linear forms. For the Zariski closure, it suffices to assume that $H_k = -\beta_1 H_1 - \cdots - \beta_{k-1} H_{k-1}$ for some general $\beta_1, \ldots, \beta_{k-1} \in \mathbb{R}$ strictly positive. So we want to write

$$\beta_i = \left(\frac{\sqrt[3]{\lambda_i}}{\sqrt[3]{\lambda_k}} \right)^2 \tag{46}$$

for some positive $\lambda_1, \ldots, \lambda_k \in \mathbb{R}$ such that $\lambda_1 + \cdots + \lambda_k = 1$. Given such λ_i, the above computations yields

$$H_1^3 + \cdots + H_k^3 = \lambda_1 L_1^3 + \cdots + \lambda_k L_k^3, \tag{47}$$

where $L_i = \frac{1}{\sqrt[3]{\lambda_i}} H_i$ for $i = 1, \ldots, k-1$ and $L_k = -\frac{\lambda_1}{\lambda_k} L_1 - \cdots - \frac{\lambda_{k-1}}{\lambda_k} L_{k-1}$, so that $\lambda_1 L_1 + \cdots + \lambda_k L_k = 0$, as wanted.

To conclude, it remains to show that Eq. (46) have a solution: these equations are equivalent to

$$(\sqrt{\beta_i})^3 = \frac{\lambda_i}{1 - \lambda_1 - \cdots - \lambda_{k-1}} \qquad \text{for } i = 1, \ldots, k-1. \tag{48}$$

Observe that the square roots are well defined since $\beta_i > 0$ for all $i = 1, \ldots, k-1$. Moreover, if $(\lambda_1, \ldots, \lambda_{k-1})$ is a solution to (48), then it is easy to see that all the λ_i must be strictly positive: indeed, since the β_i are positive, λ_i and $1 - \lambda_1 - \cdots - \lambda_{k-1}$ have the same sign. Thus, if one of the λ_i is negative, then all the λ_i are negative, but then $1 - \lambda_1 - \cdots - \lambda_{k-1} > 0$ which is absurd.

Now, setting $b_i = \sqrt{\beta_i}^3$, rewrite the equations as the linear system

$$\begin{pmatrix} 1+b_1 & b_1 & b_1 & \ldots & b_1 \\ b_2 & 1+b_2 & b_2 & \ldots & b_2 \\ b_3 & b_3 & 1+b_3 & \ldots & b_3 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b_{k-1} & b_{k-1} & b_{k-1} & \ldots & 1+b_{k-1} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \vdots \\ \lambda_{k-1} \end{pmatrix} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ \vdots \\ b_{k-1} \end{pmatrix}. \tag{49}$$

The matrix determinant lemma gives that $\det(I + b \cdot \mathbb{1}^T) = 1 + \mathbb{1}^T b = 1 + b_1 + \cdots + b_{k-1}$, which is positive since the β_i are positive. This means the system (49) has a unique solution. $\qquad \square$

Remark 5 The proof of Lemma 5 actually gives more: indeed, it shows that the image of the positive part

$$\Theta_{n,k}^{0,+} = \{((L_1, \ldots, L_k), (\lambda_1, \ldots, \lambda_k)) \in \Theta_{n,d}^0 \mid \lambda_i > 0 \text{ for all } i = 1, \ldots, k\},$$

which is the one relevant in statistics, coincides with the set of sums $\{H_1(u)^3 + \cdots + H_k(u)^3\}$, where the H_i are positively linearly dependent, meaning that there are

coefficients $\beta_1, \ldots, \beta_k > 0$ such that

$$\beta_1 H_1 + \cdots + \beta_k H_k = 0. \tag{50}$$

Remark 6 The set of sums of cubes of k dependent linear forms has a natural interpretation in terms of the projective Veronese variety: indeed consider the third Veronese embedding of $\mathbb{P}(V) = \mathbb{P}^{n-1}$:

$$v_3 \colon \mathbb{P}(V) \hookrightarrow \mathbb{P}(\mathrm{Sym}^3 V), \qquad [L] \mapsto [L^3]. \tag{51}$$

For each $(k-2)$-dimensional linear subspace $\Pi \subseteq \mathbb{P}^{n-1}$ let $\mathrm{Sec}_k(v_3(\Pi)) \subseteq \mathbb{P}(\mathrm{Sym}^3 V)$ be the k-th secant variety of its image $v_3(\Pi)$. Then, by Lemma 5, the variety $C^0_{n,k,3}$ is the affine cone over the union of these secants:

$$C^0_{n,k,3} = \mathrm{Cone}\left(\overline{\bigcup_{\Pi \subseteq \mathbb{P}^{n-1}} \mathrm{Sec}^k(v_3(\Pi))}\right). \tag{52}$$

We compute the dimension of this variety, dividing it in the cases $k \leq n+1$ and $k \geq n+1$:

Proposition 3 (i) *If $k \geq n+1$, then*

$$\dim C^0_{n,k,3} = \min\left\{kn, \binom{n+2}{3}\right\} \tag{53}$$

except in the case $n = 5, k = 7$, where $\dim C^0_{5,7,3} = 34$.
(ii) *If $k \leq n+1$, then*

$$\Delta^H_{n,4,3} = 2, \qquad \Delta^H_{n,3,3} = 2, \qquad \Delta^H_{n,2,3} = 1. \tag{54}$$

and when $k \geq 5$,

$$\Delta^H_{n,k,3} = 0. \tag{55}$$

Proof (i) Since $k \geq n+1$, Remark 6 shows that $C^0_{n,k,3}$ is the cone over the k-th secant variety $\mathrm{Sec}_k(v_3(\mathbb{P}^{n-1}))$. The dimension of this variety is computed by the Alexander–Hirschowitz theorem, so that

$$\dim C^0_{n,k,3} = \min\left\{kn, \binom{n+2}{3}\right\} \tag{56}$$

with the single exception of $n = 5, k = 7$, where the dimension is one less than the expected, hence $\dim C^0_{5,7,3} = 34$.

(ii) Since $k \leq n + 1$, Proposition 2 shows that $\Delta^H_{n,k,3} = \Delta^H_{k-1,k,3}$. Hence, for $k = 2, 3, 4$ we see directly from Table 1 that

$$\Delta^H_{3,4,3} = 2, \qquad \Delta^H_{2,3,3} = 2, \qquad \Delta^H_{1,2,3} = 1. \tag{57}$$

For $k \geq 5$ instead, we follow the proof of Proposition 2 and show that the differential of $\phi_{k-1,k,3} \colon \Theta^0_{k-1,k,3} \to \operatorname{Sym}^3 V$ at a general point is injective. For this, consider the kernel of the differential at a point $p = ((\lambda_1, \ldots, \lambda_k), (L_1, \ldots, L_k))$. It consists of elements $((\varepsilon_1, \ldots, \varepsilon_k), (H_1, \ldots, H_k)) \in \mathbb{R}^k \times V^k$ such that $\varepsilon_1 + \cdots + \varepsilon_k = 0$, $\varepsilon_1 L_1 + \cdots + \varepsilon_k L_k + \lambda_1 H_1 + \cdots + \lambda_k H_k = 0$ and

$$\sum_{i=1}^{k-1} \ell_i L_i^2 + 2h \left(\sum_{1 \leq i < j \leq k-1} \lambda_i \lambda_j L_i L_j \right) = 0, \tag{58}$$

where $\ell_i = \lambda_k^2 (3\lambda_i H_i + \varepsilon_i L_i) + \lambda_i^2 (3\lambda_k H_k + \varepsilon_k L_k)$ and $h = 3\lambda_k H_k + \varepsilon_k L_k$. Now choose the specific point p given by $\lambda_i = \frac{1}{k}$ for each $i = 1, \ldots, k$, $L_i = u_i$ for $i = 1, \ldots, k-1$ and $L_k = -u_1 - \cdots - u_{k-1}$. Then, the above equation becomes

$$\sum_{i=1}^{k-1} \ell_i u_i^2 + \frac{2}{k^2} \cdot h \cdot \left(\sum_{1 \leq i < j \leq k-1} u_i u_j \right) = 0. \tag{59}$$

Let us write $h = h_1 u_1 + \cdots + h_{k-1} u_{k-1}$. Then, in (59), the coefficient of $u_a u_b u_c$ is $\frac{2}{k^2}(h_a + h_b + h_c)$ for all $1 \leq a < b < c \leq k - 1$. Hence

$$h_a + h_b + h_c = 0, \qquad \text{for all } 1 \leq a < b < c \leq k - 1. \tag{60}$$

Let $1 \leq a < b < c < d \leq k - 1$ be any four distinct indices between 1 and $k - 1$. Then, the previous equations translate into the linear system

$$\begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} h_a \\ h_b \\ h_c \\ h_d \end{pmatrix} = 0. \tag{61}$$

The matrix appearing in the linear system is invertible, so $h_a = h_b = h_c = h_d = 0$. Since this holds for an arbitrary choice of four distinct indices, it follows that $h = 0$. Now, relation (59) tells us that $\sum_{i=1}^{k-1} u_i^2 \ell_i = 0$, but since $u_1^2, \ldots, u_{k-1}^2$ form a complete intersection of quadrics, they do not have linear syzygies, which implies that $\ell_i = 0$ for each i. From the definitions of ℓ_i and h, it follows that $3\lambda_i H_i + \varepsilon_i L_i = 0$ for each i but then the other two relations $\sum_i \varepsilon_i = 0$ and $\sum_i (\lambda_i H_i + \varepsilon_i L_i) = 0$ imply that $H_i = 0$, $\varepsilon_i = 0$ for all i, which is what was needed. $\qquad \square$

Now we are ready for a complete classification of defectivity when $d = 3$.

Table 1 All instances of defective varieties $\mathrm{Sec}_k^H(\mathcal{G}_{n,3})$ for $n = 1, \ldots, 7$ with $d = 3$

n	k	d	par	N	exp	dim	δ	Δ
1	1	3	2	3	2	2	0	0
1	2	3	4	3	3	3	0	1
2	2	3	8	9	8	7	1	1
2	3	3	11	9	9	9	0	2
3	2	3	13	19	13	12	1	1
3	3	3	17	19	17	15	2	2
3	4	3	21	19	19	19	0	2
4	2	3	19	34	19	18	1	1
4	3	3	24	34	24	22	2	2
4	4	3	29	34	29	27	2	2
4	5	3	34	34	34	34	0	0
5	2	3	26	55	26	25	1	1
5	3	3	32	55	32	30	2	2
5	4	3	38	55	38	36	2	2
5	5	3	44	55	44	44	0	0
5	6	3	50	55	50	50	0	0
5	7	3	56	55	55	54	1	2
6	2	3	34	83	34	33	1	1
6	3	3	41	83	41	39	2	2
6	4	3	48	83	48	46	2	2
6	5	3	55	83	55	55	0	0
6	6	3	62	83	62	62	0	0
6	7	3	69	83	69	69	0	0
6	8	3	76	83	76	75	1	1
6	9	3	83	83	83	81	2	2
7	2	3	43	119	43	42	1	1
7	3	3	51	119	51	49	2	2
7	4	3	59	119	59	57	2	2
7	5	3	67	119	67	67	0	0
7	6	3	75	119	75	75	0	0
7	7	3	83	119	83	83	0	0
7	8	3	91	119	91	91	0	0
7	9	3	99	119	99	98	1	1
7	10	3	107	119	107	105	2	2
7	11	3	115	119	115	112	3	3
7	12	3	123	119	119	119	0	4

The column 'par' denotes the number of parameters and 'exp' the expected dimension

Theorem 2 *For $d = 3$, the defect $\delta^H_{n,k,3} = 0$ for any k and n, with the following exceptions:*

- *$n \geq k$ and $k = 2$, where $\delta^H_{n,2,3} = 1$.*
- *$n \geq k$ and $k = 3, 4$, where $\delta^H_{n,k,3} = 2$.*
- *$n = 5$ and $k = 7$, where $\delta^H_{5,7,3} = 1$.*
- *$n \geq 4$ and $n + 1 < k \leq \frac{n^2+2n+6}{6}$ where $\delta^H_{n,k,3} = k - n - 1$.*
- *$n \geq 4$ and $\frac{n^2+2n+6}{6} \leq k < \frac{n^2+3n+2}{6}$ where $\delta^H_{n,k,3} = n\left(\frac{n^2+3n+2}{6} - k\right)$.*

Proof First consider the case when $n \geq k$: then Proposition 3 applies. It is straightforward to check that $\delta^H_{n,k,3} = \Delta^H_{n,k,d}$, from which the statement of the theorem follows.

For the cases where $k \geq n + 1$, start with the exceptional case $n = 5, k = 7$: Proposition 3 gives that $\dim \overline{\phi_{n,k,3}(\Theta^0_{5,7})} = 34$, and Lemma 3 yields $\Delta^H_{5,7,3} = 2$ and $\delta^H_{5,7,3} = 1$.

Now, consider the other cases: Proposition 3 gives that

$$\dim \overline{\phi_{n,k,3}(\Theta_{n,k})} = \min\left\{nk, \binom{n+2}{3}\right\} \tag{62}$$

and then Lemma 3 shows that

$$\Delta^H_{n,k,3} = (k-1)(n+1) - \min\left\{nk, \binom{n+2}{3}\right\}$$
$$= \max\left\{k - n - 1, k - n - 1 + n\left(k - \frac{n^2 + 3n + 2}{6}\right)\right\}$$

so that

$$\delta^H_{n,k,3} = \max\left\{k - n - 1, k - n - 1 + n\left(k - \frac{n^2 + 3n + 2}{6}\right)\right\}$$
$$- \max\left\{0, (n+1)\left(k - \frac{n^2 + 2n + 6}{6}\right)\right\}.$$

Suppose first that $n = 1, 2, 3$: this implies $k \geq n + 1 \geq \frac{n^2+2n+6}{6} \geq \frac{n^2+3n+2}{6}$ so that

$$\delta^H_{n,k,3} = k - n - 1 + n\left(k - \frac{n^2 + 3n + 2}{6}\right) - (n+1)\left(k - \frac{n^2 + 2n + 6}{6}\right) = 0.$$

Now, suppose that $n \geq 4$. Then $5 \leq n + 1 \leq \frac{n^2+2n+6}{6} \leq \frac{n^2+3n+2}{6}$ and there are three possibilities for k: if $k \geq \frac{n^2+3n+2}{6}$, then

$$\delta^H_{n,k,3} = k - n - 1 + n\left(k - \frac{n^2 + 3n + 2}{6}\right) - (n+1)\left(k - \frac{n^2 + 2n + 6}{6}\right) = 0.$$

If instead $\frac{n^2+2n+6}{6} \leq k < \frac{n^2+3n+2}{6}$, then

$$\delta_{n,k,3}^H = k - n - 1 - (n+1)\left(k - \frac{n^2+2n+6}{6}\right) = n\left(\frac{n^2+3n+2}{6} - k\right) \quad (63)$$

which is strictly positive. Finally, if $n + 1 \leq k < \frac{n^2+2n+6}{6}$, the defect is

$$\delta_{n,k,3}^H = k - n - 1, \quad (64)$$

which is positive if and only if $k > n + 1$. $\qquad\square$

As a consequence, identifiability can be characterized whenever $k \leq n + 1$:

Theorem 3 *Suppose $k \leq n + 1$. If $k \geq 5$ then a general homoscedastic mixture is algebraically identifiable from moments up to order 3. If instead $k = 2, 3, 4$, then a general homoscedastic mixture is algebraically identifiable from the moments up to order $d = 4$.*

Proof When $k \geq 5$ this follows immediately from Theorem 2 and Lemma 4. If instead $k = 2, 3, 4$, thanks to Proposition 2, it is enough to set $n = k - 1$ and check the first d for which we have identifiability: these are a finite number of cases that can be done by direct computation (e.g., in `Macaulay2` [9]), and we find that such a d is 4. $\qquad\square$

4.2 Mixtures with $k = 2$ Components

When $k = 2$ we characterize the rational identifiability as well. Since the case $d = 3$ is already covered, consider only $d \geq 4$.

Theorem 4 *The homoscedastic secant variety $\mathrm{Sec}_2^H(\mathcal{G}_{n,4})$ is algebraically identifiable. If $d \geq 5$, the homoscedastic secant variety $\mathrm{Sec}_2^H(\mathcal{G}_{n,d})$ is also rationally identifiable.*

Proof By Lemma 3 and Remark 2, it is enough to consider the parameter space given by $\Theta_{n,2}^0 = \{((L_1, L_2), (\lambda_1, \lambda_2)) \mid \lambda_1 + \lambda_2 = 1, \lambda_1 L_1 + \lambda_2 L_2 = 0\}$ and the map

$$\phi_{n,2,d} \colon \Theta_{n,2}^0 \to C_{n,2,d}^0 \subseteq \mathbb{A}_{n,d}^{K,3}. \quad (65)$$

In order to compute the general fiber of this map, note that since $d \geq 4$, it follows from Theorem 3 and its proof that the map has finite fibers. Hence, it is enough to restrict a general fiber to the open subset $\lambda_2 \neq 0$. There we may assume $L_2 = -\frac{\lambda_1}{\lambda_2}L_1 = -\frac{\lambda_1}{1-\lambda_1}L_1$. We thus compute the fibers of the induced map

$$F_{n,2,d} \colon V \times (\mathbb{R} \setminus \{1\}) \to \mathbb{A}_{n,d}^{K,3}, \qquad (L, \lambda) \mapsto \phi_{n,2,d}\left((\lambda, 1 - \lambda), \left(L, -\frac{\lambda}{1-\lambda}L\right)\right).$$

In explicit terms, this map is given by the terms from degree 3 to degree d of the logarithm $\log(\lambda e^L + (1 - \lambda)e^{-\frac{\lambda}{\lambda-1}L})$. A computation shows that the first terms are:

$$\log(\lambda e^L + (1 - \lambda)e^{-\frac{\lambda}{\lambda-1}L}) = f_3(\lambda)L^3 + f_4(\lambda)L^4 + f_5(\lambda)L^5 + \dots$$

$$f_3(\lambda) = \frac{\lambda(1 - \lambda)(1 - 2\lambda)}{6(1 - \lambda)^3}, \qquad f_4(\lambda) = \frac{\lambda(1 - \lambda)(1 - 6\lambda(1 - \lambda))}{24(1 - \lambda)^4},$$

$$f_5(\lambda) = \frac{\lambda(1 - \lambda)(1 - 2\lambda)(1 - 12\lambda(1 - \lambda))}{120(1 - \lambda)^5}. \tag{66}$$

Now suppose that $d = 4$, and let $L \in V$ and $\lambda \in \mathbb{R} \setminus \{1\}$ be general elements. In fact, it is enough to assume $L \neq 0$ and $\lambda \neq 0, 1, \frac{1}{2}$, so that $\kappa_3 = f_3(\lambda)L^3 \neq 0$. In order to compute the fiber of the point $(\kappa_3, \kappa_4) = F_{n,2,4}(L, \lambda)$, first observe that $\kappa_3 = f_3(\lambda_0)L_0^3 = (\sqrt[3]{f_3(\lambda_0)}L_0)^3$ and that the polynomial $L_0 := \sqrt[3]{f_3(\lambda)}L$ can be computed explicitly: from the expression

$$\kappa_3 = \kappa_{300..0}u_1^3 + \kappa_{030..0}u_2^3 + \cdots + \kappa_{00..03}u_n^3 + (\text{ terms with mixed monomials })$$

then one obtains

$$L_0 = \sqrt[3]{\kappa_{300..0}} \cdot u_1 + \sqrt[3]{\kappa_{030..0}} \cdot u_2 + \ldots + \sqrt[3]{\kappa_{00..03}} \cdot u_n. \tag{67}$$

In particular, $L = f_3(\lambda)^{-\frac{1}{3}}L_0$, so that the equation $\kappa_4 = f_4(\lambda)L^4$ translates into $\frac{f_4(\lambda)}{f_3(\lambda)^{\frac{4}{3}}} = \frac{\kappa_4}{L_0^4}$. Observe that $a := \frac{\kappa_4}{L_0^4}$ is a constant that can be computed explicitly by comparing a single nonzero coefficient of L_0^4 with the corresponding coefficient of κ_4: for example, if $\sqrt[3]{\kappa_{300..0}} \neq 0$, then

$$a = \frac{\kappa_{400..0}}{(\sqrt[3]{\kappa_{300..0}})^4}. \tag{68}$$

Now, the equation $\frac{f_4(\lambda)}{f_3(\lambda)^{\frac{4}{3}}} = a$ is equivalent to $\frac{f_4(\lambda)^3}{f_3(\lambda)^4} = a^3$, or more explicitly

$$\frac{3}{32} \cdot \frac{(1 - 6\lambda(1 - \lambda))^3}{\lambda(1 - \lambda)(1 - 4\lambda(1 - \lambda))^2} = a^3. \tag{69}$$

Note that this expression is invariant under exchanging λ with $1 - \lambda$, as is expected from the symmetry of the situation. Hence, set $\gamma := \lambda(1 - \lambda)$ and rewrite this expression as

$$\frac{3}{32} \cdot \frac{(1 - 6\gamma)^3}{\gamma(1 - 4\gamma)^2} = a^3. \tag{70}$$

This is a cubic equation with three possible solutions for γ, which means there is no rational identifiability. In order to get such, consider also the cumulants κ_5 of order 5: this adds the data κ_5 and the condition $\kappa_5 = f_5(\lambda)L^5$. In the above notation $L = f_3(\lambda)^{-\frac{1}{3}}L_0$, so that the condition $\kappa_5 = f_5(\lambda)L^5$ becomes $\frac{f_5(\lambda)}{f_3(\lambda)^{\frac{5}{3}}} = \frac{\kappa_5}{L_0^5}$. As before, we see that $b := \frac{\kappa_5}{L_0^5}$ is a constant that can be computed explicitly by comparing a

single nonzero coefficient of L_0^5 with the corresponding coefficient of κ_5: for example, if $\sqrt[3]{\kappa_{300..0}} \neq 0$, then

$$b = \frac{\kappa_{500..0}}{(\sqrt[3]{\kappa_{300..0}})^5}. \tag{71}$$

Now, the equation $\frac{f_5(\lambda)}{f_3(\lambda)^{\frac{5}{3}}} = a$ is equivalent to $\frac{f_5(\lambda)^3}{f_3(\lambda)^5} = b^3$, or more explicitly, as above, with the substitution $\gamma = \lambda(1 - \lambda)$,

$$\frac{15}{128} \cdot \frac{(1 - 6\gamma)^5}{\gamma(1 - \gamma)^3(1 - 12\gamma)} = b^3. \tag{72}$$

Hence, rational identifiability is obtained if the two Eqs. (70) and (72) have a unique common solution γ. This means that the map $\mathbb{R} \dashrightarrow \mathbb{R}^2$, $\gamma \mapsto (g(\gamma), h(\gamma))$ is generically injective. This map extends to $\mathbb{R} \to \mathbb{P}^2$ via

$$\left[\frac{3}{32}(1 - 6\gamma)^3(1 - \gamma)^3(1 - 12\gamma), \frac{15}{128}(1 - 6\gamma)^5(1 - 4\gamma)^2, \gamma(1 - 4\gamma)(1 - \gamma)^3(1 - 12\gamma) \right],$$

i.e., a map defined by polynomials of degree 7. It is generically injective if and only if the closure of its image is a plane curve of degree 7. This can be verified with `Macaulay2` [9]: the resulting curve is given by the equation

$$849346560x^5y^2 - 679477248x^4y^3 - 29491200x^5yz + 2674483200x^4y^2z - 2439217152x^3y^3z$$
$$+256000x^5z^2 + 79744000x^4yz^2 + 2415168000x^3y^2z^2 - 2616192000x^2y^3z^2$$
$$+499500000x^2y^2z^3 - 406500000xy^3z^3 + 474609375y^3z^4 = 0.$$

$\square$

Even though there is no rational identifiability above when $d = 4$, it is worth noting that in a purely statistical setting, γ can be recovered uniquely, as seen below.

Corollary 1 *For $k = 2$, the statistical mixture parameters can be recovered uniquely with moments up to order $d = 4$.*

Proof This is equivalent to saying that the Eq. (70) has a unique statistically relevant solution in $\gamma = \lambda(1 - \lambda)$. Note that since $\lambda \in (0, 1) \setminus \{\frac{1}{2}\}$, we have that $\gamma \in (0, \frac{1}{4})$. Consider the real valued function coming from (70):

$$a(\gamma) = \frac{\sqrt[3]{3}(1 - 6\gamma)}{2\sqrt[3]{4\gamma(1 - 4\gamma)^2}}. \tag{73}$$

Its derivative, $a'(\gamma) = -\frac{1}{2\sqrt[3]{36\gamma(1-4\gamma)}\sqrt[3]{4\gamma(1-4\gamma)^2}}$, is always negative for $0 < \gamma < \frac{1}{4}$ so that the function $a(\gamma)$ is strictly decreasing and, in particular, injective in this statistically meaningful interval (Fig. 1).
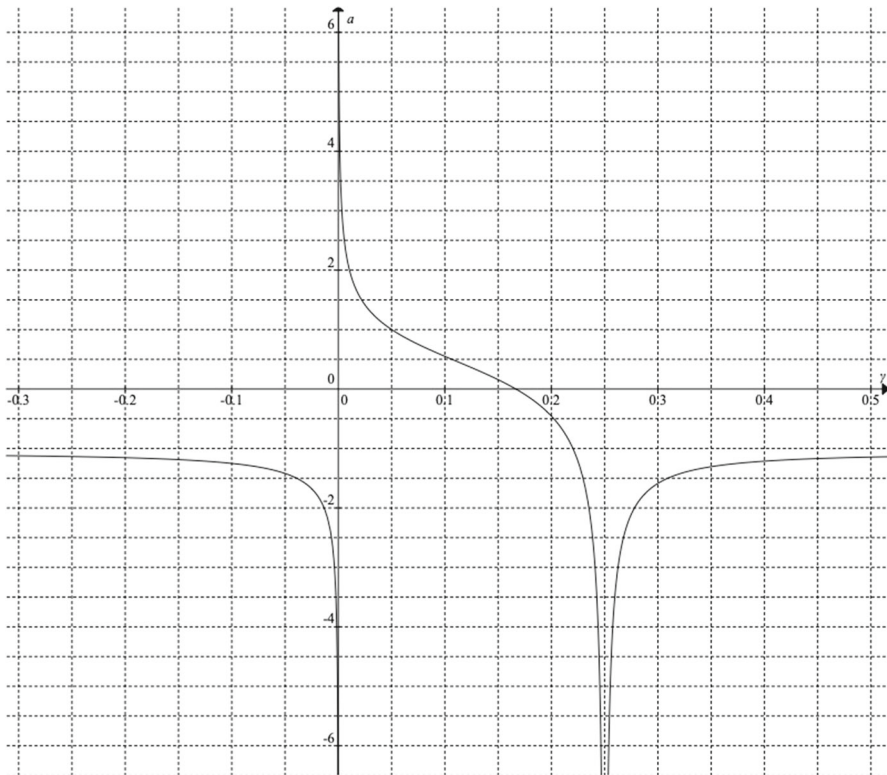
Fig. 1 Plot of the real-valued function $a(\gamma)$ in (73)

The corresponding inverse is given by the cubic equation in γ

$$(256a^3 + 324)\gamma^3 - (128a^3 + 162)\gamma^2 + (16a^3 + 27)\gamma - \frac{3}{2} = 0. \qquad (74)$$

The discriminant of (74) is $\Delta = -3072a^6(64a^3 + 81)$. It is zero precisely when $a = -\frac{3\sqrt[3]{3}}{4}$, which corresponds to the horizontal asymptote of a. If $a < -\frac{3\sqrt[3]{3}}{4}$, there are 3 real solutions, but one is negative and the other one is larger than $\frac{1}{4}$. The remaining solution is also the unique real solution when $a > -\frac{3\sqrt[3]{3}}{4}$, given explicitly by

$$\gamma = \frac{4a^3}{3\eta} + \frac{\eta}{3(64a^3 + 81)} + \frac{1}{6}, \qquad (75)$$

$$\eta = (-4096a^9 - 10368a^6 - 6561a^3 + 9\sqrt{262144a^{15} + 995328a^{12} + 1259712a^9 + 531441a^6})^{\frac{1}{3}}.$$

□

This proof gives an explicit algorithm to recover the parameters of a homoscedastic mixture of two Gaussians from the cumulants up to order four.

Algorithm 1: Recovery of parameters for a homoscedastic mixture of two Gaussians.

Data: Data coming from a homoscedastic mixture of two Gaussian distributions.
Result: The parameters $\lambda_1, \lambda_2, \mu_1, \mu_2, \Sigma$ of the mixture.
begin

> Estimate the mean vector κ_1;
> Estimate the covariance matrix κ_2;
> Estimate the principal third cumulants $\kappa_{300..0}, \kappa_{030..0}, \ldots, \kappa_{00..03}$;
> For one of the principal third cumulant that is nonzero, estimate the corresponding fourth cumulant: in the following, we assume that $\kappa_{300..0} \neq 0$, so that we estimate $\kappa_{400..0}$.;
> Compute $a = \frac{\kappa_{400..0}}{(\sqrt[3]{\kappa_{300..0}})^4}$.;
> Compute γ as in (75);
> Compute the two solutions λ_1, λ_2 of $\lambda(1-\lambda) = \gamma$;
> Compute $\mu_1' = f_3(\lambda_1)^{-\frac{1}{3}}(\sqrt[3]{\kappa_{300..0}}, \sqrt[3]{\kappa_{030..0}}, \ldots, \sqrt[3]{\kappa_{00..03}})$ and $\mu_2' = \frac{\lambda_1}{\lambda_2}\mu_1'$.;
> Compute $\mu_1 = \mu_1' + \kappa_1$ and $\mu_2 = \mu_2' + \kappa_1$;
> Compute $\Sigma = 2(\kappa_2 - (\lambda_1\mu_1 + \lambda_2\mu_2)^t(\lambda_1\mu_1 + \lambda_2\mu_2))$;

end

Observe that this algorithm needs all the cumulants of order one, all the cumulants of order two, n cumulants of order three, and one cumulant of order four. Hence, it needs in total $n + \frac{n(n+1)}{2} + n + 1$ cumulants.

Remark 7 We have seen in Remark 6 that $\mathrm{Sec}_2^H(\mathcal{G}_{n,d})$ in cumulant coordinates is a cone over $C_{n,2,d}^0 \subseteq \mathbb{A}_{n,d}^{K,3}$. Up to taking the Zariski closure, the proof of Theorem 4 shows that $C_{n,2,d}^0$ is the image of the map

$$F_{n,2,d} : V \times \mathbb{R} \setminus \{1\} \to \mathbb{A}_{n,d}^{K,3}, \quad (L, \lambda) \mapsto f_3(\lambda)L^3 + f_4(\lambda)L^4 + f_5(\lambda)L^5 + f_6(\lambda)L^6 + \ldots$$

For λ constant we get a projected d-th Veronese variety of V. If instead L is constant, then we get a rational curve given by a linear combination of $(f_3(\lambda), f_4(\lambda), \ldots, f_d(\lambda))$.

4.3 The Univariate Case $n = 1$

We use the standard notation σ^2 for the variance $\Sigma = (\sigma_{11})$ when $n = 1$.

For $n = 1$, the moment variety $\mathrm{Sec}_k^H(\mathcal{G}_{1,d})$ is never defective. The moment map

$$M_{1,k,2k} : \Theta_{1,k}^H \to \mathbb{A}_{1,2k}^M$$

is finite to one. In the statistics literature, it is known that in the case of homoscedastic secants, one may recover mixture parameters from given moments (i.e., compute the

fiber of the map above), with an algorithm closely related to the well-known Prony's method [20]. This procedure was introduced by Lindsay as an application of *moment matrices* [15] and we briefly recall the algorithm here.

First, how does one recover the locations μ_i and weights λ_i of the k components of a Dirac mixture from $2k-1$ moments? This is known as the quadrature rule and it works as follows. Given the moment sequence $m = (m_1, m_2, \ldots, m_{2k-1})$ one considers the polynomial resulting from the following $(k+1) \times (k+1)$ determinant

$$P_k(t) = \det \begin{pmatrix} 1 & m_1 & \ldots & m_{k-1} & 1 \\ m_1 & m_2 & \ldots & m_k & t \\ \vdots & & & \vdots & \vdots \\ m_k & m_{k+1} & \ldots & m_{2k-1} & t^k \end{pmatrix}. \tag{76}$$

The k roots $\mu_1, \mu_2, \ldots, \mu_k$ of $P_k(t)$ are precisely the sought locations. This follows since the equations of the secant varieties of the rational normal curve are classically known to be given by the minors of the moment matrices. For a modern reference see [14].

Once the locations are known, the weights λ_i are found by solving the $k \times k$ Vandermonde linear system

$$\begin{pmatrix} 1 & 1 & \ldots & 1 \\ \mu_1 & \mu_2 & \ldots & \mu_k \\ \vdots & & & \vdots \\ \mu_1^{k-1} & \mu_2^{k-1} & \ldots & \mu_k^{k-1} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_k \end{pmatrix} = \begin{pmatrix} 1 \\ m_1 \\ \vdots \\ m_{k-1} \end{pmatrix}. \tag{77}$$

Back to the Gaussian case, if we knew the value of the common variance σ^2, we can reduce to the above instance. In terms of the Gaussian moment generating function:

$$e^{-\frac{1}{2}\sigma^2 u^2} M_X(u) = e^{\mu u}. \tag{78}$$

Hence, the Dirac moments $\tilde{m}$ on the right hand side are linear combinations of the Gaussian moments m. Explicitly, for $1 \le j \le 2k - 1$

$$\tilde{m}_j(\sigma) = \sum_{i=0}^{\lfloor j/2 \rfloor} \frac{j!}{(-2)^i i!(j - 2i)!} m_{j-2i} \sigma^{2i}. \tag{79}$$

Applying the quadrature rule to the vector $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_{2k-1})$ would allow us to obtain the means $\mu_1, \mu_2, \ldots, \mu_k$.

However, σ is unknown. To find an estimate for σ we consider the first $2k$ moments $m = (m_1, m_2, \ldots, m_{2k})$. If $\tilde{m} = (\tilde{m}_1, \tilde{m}_2, \ldots, \tilde{m}_{2k})$ comes from a mixture of k Dirac

measures, then

$$D_k = \det \begin{pmatrix} 1 & \tilde{m}_1 & \ldots & \tilde{m}_{k-1} & \tilde{m}_k \\ \tilde{m}_1 & \tilde{m}_2 & \ldots & \tilde{m}_k & \tilde{m}_{k+1} \\ \vdots & & & \vdots & \vdots \\ \tilde{m}_k & \tilde{m}_{k+1} & \ldots & \tilde{m}_{2k-1} & \tilde{m}_{2k} \end{pmatrix} = 0. \tag{80}$$

One thus treats σ as a variable and substitutes expressions (79) into (80). This results in a polynomial $D_k(\sigma)$ of degree $\binom{k+1}{2}$ in σ^2 and the estimator $\hat{\sigma}^2$ is obtained as its smallest non-negative root [15, Theorem 5B]. So the algebraic degree for estimating σ^2 is $\binom{k+1}{2}$. With σ^2 specified, one proceeds as above.

More generally, the discussion under (28) shows that the moment variety $\text{Sec}_k^H(\mathcal{G}_{1,d})$ with $k \leq d/2$ is a union

$$\text{Sec}_k^H(\mathcal{G}_{1,d}) = \bigcup_\sigma \text{Sec}_k(V_{1,d}^\sigma),$$

where $V_{1,d}^\sigma$ is the translation of the moment curve $V_{1,d}$ by the variance σ^2 as defined by the Gaussian moments. The secant variety $\text{Sec}_k(V_{1,d}^\sigma)$ is defined for each σ by the $(k+1) \times (k+1)$ minors of

$$M_{k,d} = \begin{pmatrix} 1 & \tilde{m}_1 & \ldots & \tilde{m}_{d-k-1} & \tilde{m}_{d-k} \\ \tilde{m}_1 & \tilde{m}_2 & \ldots & \tilde{m}_{d-k} & \tilde{m}_{d-k+1} \\ \vdots & & & \vdots & \vdots \\ \tilde{m}_k & \tilde{m}_{k+1} & \ldots & \tilde{m}_{d-1} & \tilde{m}_d \end{pmatrix}. \tag{81}$$

As soon as the k-th secant variety of a smooth curve is not linear, the curve can be recovered as the singular locus of highest multiplicity in the secant variety. Therefore, since curves $V_{1,d}^\sigma$ are distinct, their k-th secant varieties are distinct as well, as long as the latter are not linear. In particular, since the variety $\text{Sec}_k(V_{1,d}^\sigma)$ has dimension $2k - 1$, it follows that the union $\text{Sec}_k^H(\mathcal{G}_{1,d})$ has dimension $2k$. Given the moments m_i up to degree d of a point on a homoscedastic k-secant, the $(k+1) \times (k+1)$ minors of $M_{k,d}$ are polynomials in σ^2 with a zero at the common variance. Given the variance, the means can be inferred as above.

When $d = 2k+1$, then the variety $\text{Sec}_k^H(\mathcal{G}_{1,d}) \subset \mathbb{A}_{1,2k+1}^M$ is a hypersurface, defined by the resultant of $(k+1)$-minors of $M_{k,d}$, the polynomial obtained by elimination of σ^2 in the ideal defined by the $(k+1) \times (k+1)$ minors. Denote this polynomial by P_{2k+1}. It is a polynomial in $m_1, \ldots, m_{2k+1}$ (or $\kappa_3, \kappa_4, \ldots, \kappa_{2k+1}$). For example,

$$P_3 = \kappa_3 = 2m_1^3 - 3m_1 m_2 + m_3,$$
$$P_5 = 108\kappa_3^6 - 32\kappa_3^2\kappa_4^3 + 36\kappa_3^3\kappa_4\kappa_5 - \kappa_4^2\kappa_5^2 + \kappa_3\kappa_5^3$$

Proposition 4 *The polynomial P_{2k+1} is homogeneous of total degree*

$$\binom{k+2}{2}\binom{k+1}{2}$$

in the multigraded weights $\deg m_i = \deg \kappa_i = i$.

Proof Let

$$\mathbb{A} = \mathbb{A}_{1,2k+1}^M \times \mathbb{A}^1$$

where σ is the last coordinate, and consider the projective closure $\mathbb{P}$ of $\mathbb{A}$. Then, the matrix (81) defines a map between vector bundles E and F on $\mathbb{A}$. The vector bundles E and F and the map extends to $\mathbb{P}$; E extends to a sum of line bundles $\tilde{E} = \mathcal{O}_\mathbb{P} \oplus \mathcal{O}_\mathbb{P}(-1) \oplus \ldots, \mathcal{O}_\mathbb{P}(-k)$, while F extends to a sum of line bundles $\tilde{F} = \mathcal{O}_\mathbb{P} \oplus \mathcal{O}_\mathbb{P}(1) \oplus \ldots, \mathcal{O}_\mathbb{P}(k+1)$. By the Thom–Porteous formula, [8, Theorem 14.4], the degree in $\mathbb{P}$ of the rank k locus of the map is given by the Chern class

$$c_2(\tilde{F} - \tilde{E}) = 2\binom{k+2}{2}\binom{k+1}{2}$$

since the Chern polynomials of $\tilde{E}$ and $\tilde{F}$ in are

$$c(\tilde{E}) = (1 - t)(1 - 2t)\ldots(1 - kt)$$

and

$$c(\tilde{F}) = (1 + t)(1 + 2t)\ldots(1 + (k+1)t).$$

This rank k locus has codimension 2 and its intersection with $\mathbb{A}$ is projected to the hypersurface defined by P_{2k+1} in $\mathbb{A}_{1,2k+1}^M$. The coordinate σ appears only in even degree in the equations defining the rank k locus, so the projection to $\mathbb{A}_{1,2k+1}^M$ is 2 : 1, so the degree of P_{2k+1} is half the degree of the rank k locus. □

Question 1 It would be interesting to understand better the structure of the polynomials P_{2k+1}, e.g., is there a closed form expression for all k?

If P_{2k+1} vanishes on a set $(m_1, \ldots, m_{2k+1})$ of moments, and P_{2l+1} does not vanish on $(m_1, \ldots, m_{2l+1})$ for any $l < k$, then the moments lie on a homoscedastic k-secant but not on any l secant for $l < k$. Therefore the polynomials P_{2k+1} may be used to estimate the number of components in a homoscedastic Gaussian mixture (compare to the rank test proposed in [15, Section 3.1] for the known variance case).

5 Conclusion

We have completely classified all defective cases for the moment varieties associated with homoscedastic Gaussian mixtures whenever $k < n + 1$, $d = 3$, $k = 2$ or $n = 1$. The question concerning a complete classification for all n, d, k remains open, although our computations did not reveal any further defective examples.

Our identifiability results also cover special structures in the covariance matrix, by Remark 2. For example, a common mixture submodel involves isotropic Gaussians, which means that the covariance matrix is a scalar multiple of the identity, $\Sigma = \sigma I$. The k-means algorithm used in clustering can be interpreted as parameter estimation for a homoscedastic isotropic mixture of Gaussians. In [10], Hsu and Kakade consider the learning of mixtures of isotropic Gaussians from the moments up to order $d = 3$ when $k \leq n + 1$. They prove identifiability for the homoscedastic isotropic submodel (see [6, Theorem 3.2]), and in order to solve the moment equations, they find orthogonal decompositions of the second and third order moment tensors.

On the other hand, in [17] Lindsay and Basak proposed a 'fast consistent' method of moments for homoscedastic Gaussian mixtures in the multivariate case, based on a 'primary axis' to which the one-dimensional case presented in Sect. 4.3 is applied. This means that the method uses some moments of order $2k$. Knowing that in some cases there are explicit equations for secant varieties of higher dimensional Veronese varieties [14], an alternative method with minimal order based on these should be possible.

Finally, a similar approach can be made to study moment varieties of homoscedastic mixtures of other location families. In the case of Example 4, we saw that Gaussian moments and Laplacian moments coincide up to $d = 3$. This means that Theorem 2 applies verbatim to homoscedastic mixtures of Laplace distributions.

References

1. Agostini, D., Améndola, C.: Discrete Gaussian distributions via theta functions. SIAM Journal on Applied Algebra and Geometry **3**(1), 1–30 (2019)
2. Alexander, J., Hirschowitz, A.: Polynomial interpolation in several variables. Journal of Algebraic Geometry **4**(2), 201–222 (1995)
3. Améndola, C.: Algebraic statistics of Gaussian mixtures. Ph.D. thesis, Technische Universität Berlin (2017). https://doi.org/10.14279/depositonce-6557

4. Améndola, C., Faugère, J.C., Sturmfels, B.: Moment varieties of Gaussian mixtures. Journal of Algebraic Statistics **7**, 14–28 (2016)
5. Améndola, C., Ranestad, K., Sturmfels, B.: Algebraic identifiability of Gaussian mixtures. International Mathematics Research Notices (2017)
6. Anandkumar, A., Ge, R., Hsu, D.J., Kakade, S.M., Telgarsky, M.: Tensor decompositions for learning latent variable models. Journal of Machine Learning Research **15**(1), 2773–2832 (2014)
7. Brambilla, M.C., Ottaviani, G.: On the Alexander–Hirschowitz theorem. Journal of Pure and Applied Algebra **212**(5), 1229–1251 (2008)
8. Fulton, W.: Intersection theory, vol. 2. Springer, UK (2013)
9. Grayson, D.R., Stillman, M.E.: Macaulay 2, a software system for research in algebraic geometry (2002)
10. Hsu, D., Kakade, S.M.: Learning mixtures of spherical Gaussians: moment methods and spectral decompositions. In: Proceedings of the 4th conference on Innovations in Theoretical Computer Science, pp. 11–20. ACM (2013)
11. Kohn, K., Shapiro, B., Sturmfels, B.: Moment varieties of measures on polytopes. Annali della Scuola Normale Superiore di Pisa (Classe Scienze), Serie V (2019). https://doi.org/10.2422/2036-2145.201808003
12. Kotz, S., Kozubowski, T., Podgorski, K.: The Laplace distribution and generalizations: a revisit with applications to communications, economics, engineering, and finance. Springer, UK (2012)
13. Koutsoumpelias, A.G., Wageringel, M.: Moment ideals of local Dirac mixtures. SIAM Journal on Applied Algebra and Geometry **4**(1), 1–27 (2020)
14. Landsberg, J.M., Ottaviani, G.: Equations for secant varieties of Veronese and other varieties. Annali di Matematica Pura ed Applicata **192**(4), 569–606 (2013)
15. Lindsay, B.G.: Moment matrices: applications in mixtures. The Annals of Statistics pp. 722–740 (1989)
16. Lindsay, B.G.: Method of moments. Wiley StatsRef: Statistics Reference Online (2014)
17. Lindsay, B.G., Basak, P.: Multivariate normal mixtures: a fast consistent method of moments. Journal of the American Statistical Association **88**(422), 468–476 (1993)
18. Pearson, K.: Contributions to the mathematical theory of evolution. Philosophical Transactions of the Royal Society of London. A **185**, 71–110 (1894)
19. Sullivant, S.: Algebraic Statistics, *Graduate Studies in Mathematics*, vol. 194. American Mathematical Society, Providence, RI (2018)
20. Weiss, L., McDonough, R.: Prony's method, z-transforms, and Padé approximation. Siam Review **5**(2), 145–149 (1963)
21. Wu, Y., Yang, P.: Optimal estimation of Gaussian mixtures via denoised method of moments. The Annals of Statistics, to appear. arXiv preprint arXiv:1807.07237 (2018)