PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 147, Number 7, July 2019, Pages 3021–3038
https://doi.org/10.1090/proc/14499
Article electronically published on March 21, 2019

THE MULTIDIMENSIONAL TRUNCATED MOMENT PROBLEM: GAUSSIAN AND LOG-NORMAL MIXTURES, THEIR CARATHÉODORY NUMBERS, AND SET OF ATOMS

PHILIPP J. DI DIO

(Communicated by David Levin)

Abstract. We study truncated moment sequences of distribution mixtures, especially from Gaussian and log-normal distributions and their Carathéodory numbers. For $\mathsf{A} = \{a_1, \ldots, a_m\}$ continuous (sufficiently differentiable) functions on $\mathbb{R}^n$ we give a general upper bound of $m-1$ and a general lower bound of $\left\lceil \frac{2m}{(n+1)(n+2)} \right\rceil$. For polynomials of degree at most d in n variables we find that the number of Gaussian and log-normal mixtures is bounded by the Carathéodory numbers in [J. Math. Anal. Appl. 461 (2018), pp. 1606–1638]. Therefore, for univariate polynomials $\{1, x, \ldots, x^d\}$ at most $\left\lceil \frac{d+1}{2} \right\rceil$ distributions are needed. For bivariate polynomials of degree at most $2d-1$ we find that $\frac{3d(d-1)}{2} + 1$ Gaussian distributions are sufficient. We also treat polynomial systems with gaps and find, e.g., that for $\{1, x^2, x^3, x^5, x^6\}$ three Gaussian distributions are enough for almost all truncated moment sequences. For log-normal distributions the number is bounded by half of the moment number. We give an example of continuous functions where more Gaussian distributions are needed than Dirac delta measures. We show that any inner truncated moment sequence has a mixture which contains any given distribution.

1. Introduction

In many applications, the distribution is a linear combination of simple distributions such as Gaussian distributions

$$(1)\qquad g_{\xi,\sigma}(x) := \frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot e^{-\frac{(x-\xi)^2}{2\sigma^2}} \quad \text{with} \quad \xi \in \mathbb{R},\ \sigma > 0,$$

or log-normal distributions

$$(2)\qquad l_{\xi,\sigma}(x) := \begin{cases} \frac{1}{\sqrt{2\pi}\cdot\sigma x}\cdot e^{-\frac{(\log x - \log \xi)^2}{2\sigma^2}} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \quad \text{with} \quad \xi, \sigma \in (0, \infty).$$

E.g., in the seminal paper of K. Pearson he investigates the distribution of the breadth of the foreheads of Naples crabs and the length of the carapace of prawns

Received by the editors May 8, 2018, and, in revised form, July 19, 2018, and October 13, 2018.

2010 *Mathematics Subject Classification.* Primary 44A60, 14P10.

Key words and phrases. Truncated moment problem, Carathéodory number, Gaussian, log-normal, finite mixture models, moment method.

The author was supported by the Deutsche Forschungsgemeinschaft (SCHM1009/6-1).

[Pea94]. Since the data did not fit a single Gaussian distribution, he assumed that the distribution comes from a linear combination of two Gaussian distributions

$$\frac{c_1}{\sigma_1\sqrt{2\pi}}\cdot e^{-\frac{(x-x_1)^2}{2\sigma_1^2}} + \frac{c_2}{\sigma_2\sqrt{2\pi}}\cdot e^{-\frac{(x-x_2)^2}{2\sigma_2^2}}. \tag{3}$$

To determine c_1, c_2, σ_1, σ_2, x_1, and x_2 he calculated the first five moments of (3) (all are polynomials in $c_1,\dots,x_2$) and after algebraic manipulations got a polynomial of degree 9. The zeros of this polynomial are the solution of fitting (3) to the crab data. This method is now well known by the name *method of moments*; see, e.g., [TSM85].

Another frequent distribution is the log-normal distribution (2). It appears, e.g., in the study of option pricings in financial mathematics [Sto16], especially in the Black–Scholes model by Black, Scholes [BS73], and Merton [Mer73]. In that model it is found that the option pricing is given by

$$x\cdot g_{0,1}(d_1) - c\cdot e^{r(t-t^*)}\cdot g_{0,1}(d_2) \tag{4}$$

with

$$d_1 = \frac{\log(x/c) + (r+\frac{v^2}{2})(t^*-t)}{v\sqrt{t^*-t}} \quad \text{and} \quad d_2 = \frac{\log(x/c) + (r-\frac{v^2}{2})(t^*-t)}{v\sqrt{t^*-t}},$$

where t is the time variable and x, c, t^*, r, and v are parameters of the option/model. Despite the fact that in the Black–Scholes model the linear combination (4) depends on several parameters and is only related to the log-normal distribution, the log-normal distribution (2) is frequently used and one of the most important distributions in financial engineering [Sto16].

In the following article we treat the problem of mixtures of densities very generally, but we also derive more detailed results for the Gaussian (1) and log-normal distribution (2) because of their importance. We use the following *general setting*:

(a) $\delta_{\xi,\sigma}$ are probability measures on a (topological) space $\mathcal{X}$ with parameters $\xi\in\mathcal{X}$ and $\sigma\in\Sigma$, Σ is the set of parameters (variance; in a larger metric space).

(b) $\mathsf{A} = \{a_1,\dots,a_m\}$ is a set of linearly independent (real valued) continuous functions on the space $\mathcal{X}$ s.t.

$$\left|\int_{x\in\mathcal{X}} a_i(x)\ \mathrm{d}\delta_{\xi,\sigma}(x)\right| < \infty \quad \forall\xi\in\mathcal{X},\ \sigma\in\Sigma.$$

(c) There exists a $\sigma_0\in\overline{\Sigma}$ (closure of Σ) such that

$$\lim_{\Sigma\ni\sigma\to\sigma_0}\int_{x\in\mathcal{X}} a_i(x)\ \mathrm{d}\delta_{\xi,\sigma}(x) = a_i(\xi) \quad \forall\xi\in\mathcal{X},\ i=1,\dots,m.$$

(d) If the integral $s_i := \int_{\mathcal{X}} a_i(x)\ \mathrm{d}\mu(x)$ exists it is called an ith (or a_i-)moment of the measure μ.

The name moment problem comes from $\mathsf{A} = \{1,x,x^2,\dots,x^d\}$, i.e., the (classical) moments are $\int_{\mathcal{X}} x^i\ \mathrm{d}\mu(x)$, while the general moments are $\int_{\mathcal{X}} a_i(x)\ \mathrm{d}\mu(x)$. Truncated means that only finitely many moments of μ are known (A is finite). Of course, since the integral is linear in the integrand, the moment problem rather depends on $\operatorname{lin}\mathsf{A}$ than on A. So we can always choose an appropriate basis A of $\operatorname{lin}\mathsf{A}$.

Example 1. For the Gaussian distributions (1) we have $\mathcal{X} = \mathbb{R}^n$ ($n \in \mathbb{N}$), $\Sigma \subset \mathbb{R}^{n\times n}$ is the set of all symmetric non-singular matrices, and $\sigma_0 = 0 \in \mathbb{R}^{n\times n}$ is the zero matrix. The *Gaussian measure* $\delta^G_{\xi,\sigma}$ is then defined by

$$\text{(5)}\quad \mathrm{d}\delta^G_{\xi,\sigma}(x) := G_{\xi,\sigma}(x)\ \mathrm{d}\lambda^n(x) \quad \text{with} \quad G_{\xi,\sigma}(x) := \frac{\exp\left(-\frac{1}{2}(x-\xi)^T\sigma^{-2}(x-\xi)\right)}{\sqrt{(2\pi)^n \det(\sigma)^2}}$$

and λ^n is the n-dimensional Lebesgue measure. $\mathsf{A} = \{a_1,\dots,a_m\} \subset C(\mathbb{R}^n,\mathbb{R})$ is a linearly independent set of continuous functions s.t. (b) holds. By continuity of the a_i's (b) holds. Then (c) holds, i.e., the Dirac delta measure δ_ξ is approximated by $\delta^G_{\xi,\sigma}$ if $\sigma \to \sigma_0 = 0$.

Example 2. Similarly, for the log-normal distribution (2) we have $\mathcal{X} = \mathbb{R}^n$ (or $\mathcal{X} = (0,\infty)^n$ with $n \in \mathbb{N}$), again $\Sigma \subset \mathbb{R}^{n\times n}$ is the set of all symmetric non-singular matrices, $\sigma_0 = 0 \in \mathbb{R}^{n\times n}$ the zero matrix. We define the *log-normal measure* $\delta^L_{\xi,\sigma}$ by

$$\text{(6)}\quad \begin{aligned} &\mathrm{d}\delta^L_{\xi,\sigma}(x) := L_{\xi,\sigma}(x)\ \mathrm{d}\lambda^n(x) \quad \text{with} \\ &L_{\xi,\sigma}(x) := \begin{cases} \frac{\exp\left(-\frac{1}{2}(\log x - \log\xi)^T\sigma^{-2}(\log x-\log\xi)\right)}{\sqrt{(2\pi)^n\det(\sigma)^2}\cdot\prod_{i=1}^n x_i} & \text{for } x_1,\dots,x_n > 0,\\ 0 & \text{else,}\end{cases}\end{aligned}$$

where $\log x := (\log x_i)_{i=1}^n$ and λ^n is again the n-dimensional Lebesgue measure. $\mathsf{A} = \{a_1,\dots,a_m\} \subset C(\mathcal{X},\mathbb{R})$ is a linearly independent set of continuous functions s.t. (b) holds. Then (c) holds, i.e., the Dirac delta measure δ_ξ is approximated by $\delta^L_{\xi,\sigma}$ if $\sigma \to \sigma_0 = 0$.

For the Gaussian (1) and the log-normal distribution (2) all moments are known and finite ($n = 1$):

$$\text{(7)}\quad \int_{\mathbb{R}} (x-\xi)^i\ \mathrm{d}\delta^G_{\xi,\sigma}(x) = \begin{cases} (i-1)!!\cdot\sigma^i & \text{for } 2|i,\\ 0 & \text{else,}\end{cases}$$

and

$$\text{(8)}\quad \int_0^\infty x^i\ \mathrm{d}\delta^L_{\xi,\sigma}(x) = \xi^i \cdot e^{\frac{i^2\sigma^2}{2}}.$$

For $n > 1$ similar formulas hold by diagonalizing σ.

We investigate *mixtures of distributions*

$$\text{(9)}\quad \sum_{i=1}^k c_i\cdot\delta_{\xi_i,\sigma_i} \qquad (c_i > 0)$$

with the moment method. In previous works and applications the number k of components is fixed and justified by the model or the data and one of the main questions is the identifiability (uniqueness/determinacy) of (9); see, e.g., [Pea94], [BS73], [TSM85], [MMR05], [PFJ06], [Sto16], [AFS16], [ABB+17], [ARS17], and the references therein. But in the present paper we want to investigate the moment cone (Section 3), the possible $\delta_{\xi,\sigma}$ appearing in a representation (9) (Section 4), and the number k of components needed to represent a given finite number of moments (Section 5).

2. Preliminaries

The theory and application of moments is rich; see, e.g., [KS53], [Ric57], [Rog58], [AK62], [Akh65], [Kem68], [KN77], [Sch91], [Mat92], [Rez92], [CF96a], [CF96b], [Sim98], [CF00], [Sch03], [FP05], [CF05], [PS08], [Mar08], [Lau09], [FN10], [CF13], [Las15], [Sch15], [Sto16], [Fia17], [IKLS17], [SdD17], [Sch17], [RS18], [dDS18a], [dDS18b], and the references therein. But in the present section we only present definitions and results needed in the following sections, especially from [dDS18a] and [dDS18b] with extensions to mixtures as presented in the introduction.

To efficiently deal with (linear combinations of) Dirac measures δ_ξ and probability measures $\delta_{\xi,\sigma}$ we introduce the following.

Definition 3. The *moment curve* s_{A} is defined by

$$s_{\mathsf{A}} : \mathcal{X} \to \mathbb{R}^m,\ x \mapsto s_{\mathsf{A}}(x) := \begin{pmatrix} a_1(x) \\ \vdots \\ a_m(x) \end{pmatrix}$$

and for $k \in \mathbb{N}$ the *moment map* is defined by

$$S_{k,\mathsf{A}} : \mathbb{R}^k_{\geq 0} \times \mathcal{X}^k \to \mathbb{R}^m,\ (C, X) \mapsto S_{k,\mathsf{A}}(C, X) := \sum_{i=1}^k c_i \cdot s_{\mathsf{A}}(x_i),$$

where $C = (c_1, \ldots, c_k)$ and $X = (x_1, \ldots, x_k)$. We denote by $\mathcal{M}_{\mathsf{A}}$ the set of all (positive) measures μ on $\mathcal{X}$ s.t. $\left|\int_{\mathcal{X}} a_i(x)\ \mathrm{d}\mu(x)\right| < \infty$ for all $i = 1, \ldots, m$.

Clearly, $s_{\mathsf{A}}(x)$ is the moment sequence of the Dirac measure δ_x and $S_{k,\mathsf{A}}(C, X)$ is the moment sequence of the measure $\mu = \sum_{i=1}^k c_i \cdot \delta_{x_i}$. This and further definitions of course depend on the choice and order of the a_i's in A. But since the integral is linear in the integrand, reordering or changing the basis A does not affect our results. We also write $\mu = (C, X)$ for a finitely atomic measure and we have $\delta_x, (C, X) \in \mathcal{M}_{\mathsf{A}}$. To deal with $\delta_{\xi,\sigma}$ we introduce the following.

Definition 4. We define

$$t_{\mathsf{A}} : \mathcal{X} \times \Sigma \to \mathbb{R}^m,\ (x, \sigma) \mapsto t_{\mathsf{A}}(x, \sigma) := \left(\int_{\mathcal{X}} a_i(y)\ \mathrm{d}\delta_{x,\sigma}(y)\right)_{i=1}^m$$

and

$$T_{k,\mathsf{A}} : \mathbb{R}^k_{\geq 0} \times \mathcal{X}^k \times \Sigma^k \to \mathbb{R}^m,\ (C, X, \bar{\sigma}) \mapsto T_{k,\mathsf{A}}(C, X, \bar{\sigma}) := \sum_{i=1}^k c_i \cdot t_{\mathsf{A}}(x_i, \sigma_i),$$

where $C = (c_1, \ldots, c_k)$, $X = (x_1, \ldots, x_k)$, and $\bar{\sigma} = (\sigma_1, \ldots, \sigma_k)$.

Clearly, $t_{\mathsf{A}}(x, \sigma)$ is the moment sequence of $\delta_{x,\sigma} \in \mathcal{M}_{\mathsf{A}}$ and $T_{k,\mathsf{A}}(C, X, \bar{\sigma})$ is the moment sequence of the mixture $\mu = (C, X, \bar{\sigma}) = \sum_{i=1}^k c_i \cdot \delta_{x_i,\sigma_i} \in \mathcal{M}_{\mathsf{A}}$. From condition (c) we get

$$\lim_{\Sigma \ni \sigma \to \sigma_0} t_{\mathsf{A}}(x, \sigma) = s_{\mathsf{A}}(x). \tag{10}$$

Definition 5. We define the *moment cone*

$$\mathcal{S}_{\mathsf{A}} := \left\{ \int_{\mathcal{X}} s_{\mathsf{A}}(x)\ \mathrm{d}\mu(x) \,\middle|\, \mu \in \mathcal{M}_{\mathsf{A}} \right\} \subseteq \mathbb{R}^m,$$

its boundary points

$$\partial^* \mathcal{S}_{\mathsf{A}} := \partial \mathcal{S}_{\mathsf{A}} \cap \mathcal{S}_{\mathsf{A}},$$

and the set

$$\mathcal{T}_{\mathsf{A}} := T_{m,\mathsf{A}}(\mathbb{R}^m_{\geq 0} \times \mathcal{X}^m \times \Sigma^m) = \operatorname{range} T_{m,\mathsf{A}}.$$

$\mathcal{T}_{\mathsf{A}}$ is the set of all moment sequences which have a mixture (9) as a representing measure with at most m components. It will turn out that $\mathcal{T}_{\mathsf{A}}$ is a convex full-dimensional cone; see Theorem 17. Of course, $\mathcal{T}_{\mathsf{A}} \subseteq \mathcal{S}_{\mathsf{A}}$ since $(C, X, \bar{\sigma}) \in \mathcal{M}_{\mathsf{A}}$ by (b). For the Dirac measures we have the following theorem due to H. Richter. See, e.g., [Sch17, Thm. 1.24] for a more recent proof.

Theorem 6 (H. Richter 1957 [Ric57, Satz 4])**.** *Let $\mathcal{X}$ be a topological space, and let $\mathsf{A} = \{a_1, \dots, a_m\}$ be a finite set of functions on $\mathcal{X}$, i.e., $\delta_x \in \mathcal{M}_{\mathsf{A}}$ for all $x \in \mathcal{X}$. Then*

$$\mathcal{S}_{\mathsf{A}} = \operatorname{range} S_{m,\mathsf{A}} = S_{m,\mathsf{A}}(\mathbb{R}^m_{\geq 0} \times \mathcal{X}^m),$$

i.e., for every $\mu \in \mathcal{M}_{\mathsf{A}}$ there is a finitely atomic measure $\mu' = (C, X) = \sum_{i=1}^k c_i \cdot \delta_{x_i}$ with the same moment sequence $\int_{\mathcal{X}} a_i(x)\ \mathrm{d}\mu(x) = \int_{\mathcal{X}} a_i(x)\ \mathrm{d}\mu'(x)$ and $k \leq m$.

By the Richter Theorem (Theorem 6) every moment sequence $s \in \mathcal{S}_{\mathsf{A}}$ has a finitely atomic representing measure and we can introduce the following number.

Definition 7. Let $s \in \mathcal{S}_{\mathsf{A}}$. We call $\mathcal{C}_{\mathsf{A}}(s)$ defined by

$$\mathcal{C}_{\mathsf{A}}(s) := \min\{k \in \mathbb{N} \mid s \in \operatorname{range} S_{k,\mathsf{A}}\}$$

the *Carathéodory number of s*. The *Carathéodory number* $\mathcal{C}_{\mathsf{A}}$ is

$$\mathcal{C}_{\mathsf{A}} := \max_{s \in \mathcal{S}_{\mathsf{A}}} \mathcal{C}_{\mathsf{A}}(s).$$

For the special case of univariate polynomials Richter also proved the following famous result.

Theorem 8 (H. Richter 1957 [Ric57, Satz 11])**.** *Let $\mathsf{A} = \{1, x, \dots, x^d\}$ on an open, half-open, or closed interval of $\mathbb{R}$ (or $\mathcal{X} = \mathbb{R}$). Then*

$$\mathcal{C}_{\mathsf{A}} = \left\lceil \frac{d+1}{2} \right\rceil.$$

In [dDS18b] we introduced the following important number.

Definition 9. Let $\mathsf{A} = \{a_1, \dots, a_m\} \subset C^1(U, \mathbb{R})$ be a linearly independent subset of C^1-functions on an open set $U \subseteq \mathbb{R}^n$. Define

$$\mathcal{N}_{\mathsf{A}} := \min\{k \in \mathbb{N} \mid DS_{k,\mathsf{A}} \text{ has full rank}\}, \tag{11}$$

where $DS_{k,\mathsf{A}}$ denotes the total derivative

$$\begin{aligned} DS_{k,\mathsf{A}} &= (\partial_{c_1} S_{k,\mathsf{A}}, \partial_{x_{1,1}} S_{k,\mathsf{A}}, \dots, \partial_{x_{1,n}} S_{k,\mathsf{A}}, \partial_{c_2} S_{k,\mathsf{A}}, \dots, \partial_{x_{k,n}} S_{k,\mathsf{A}}) \\ &= (s_{\mathsf{A}}(x_1), c_1 \partial_1 s_{\mathsf{A}}(x_1), \dots, c_1 \partial_n s_{\mathsf{A}}(x_1), s_{\mathsf{A}}(x_2), \dots, c_k \partial_n s_{\mathsf{A}}(x_k)) \end{aligned} \tag{12}$$

of $S_{k,\mathsf{A}}$.

We also proved the following general lower bound on $\mathcal{C}_{\mathsf{A}}$ using Sard's Theorem [Sar42].

Theorem 10 ([dDS18b, Thm. 27])**.** *Let* $\mathsf{A} = \{a_1, \dots, a_m\} \subset C^r(\mathcal{X}, \mathbb{R})$ *be linearly independent with* $\mathcal{X} \subseteq \mathbb{R}^n$ *and* $r > \mathcal{N}_{\mathsf{A}}(n+1) - m$*. Then*

$$\left\lceil \frac{m}{n+1} \right\rceil \leq \mathcal{N}_{\mathsf{A}} \leq \mathcal{C}_{\mathsf{A}} \tag{13}$$

and the set of moment sequences s *with* $\mathcal{C}_{\mathsf{A}}(s) < \mathcal{N}_{\mathsf{A}}$ *has* m*-dimensional Lebesgue measure zero in* $\mathbb{R}^m$*.*

Remark 11. Instead of $\mathcal{X}$ being an open subset of $\mathbb{R}^n$, we could extend Definition 9 and Theorem 10 to (differentiable) manifolds $\mathcal{X}$. By choosing a chart $\varphi : U \subseteq \mathbb{R}^n \to \mathcal{X}$ of the manifold, U open, we have again the previous definition and theorem for $\mathsf{A}\circ\varphi = \{a_i \circ \varphi \,|\, i = 1, \dots, m\}$. It therefore suffices to treat $\mathcal{X} \subseteq \mathbb{R}^n$ open or $\mathcal{X} = \mathbb{R}^n$.

For upper bounds we proved an $(m-1)$-theorem, which we will tighten here.

Theorem 12 (An extension of [dDS18b, Thm. 13])**.** *Let* A *and* $\mathcal{X}$ *s.t. there exists an* $e \in \operatorname{lin} \mathsf{A}$ *with* $e(x) > 0$ *for all* $x \in \mathcal{X}$ *and* $\operatorname{range} s_{\mathsf{A}} \cdot \|s_{\mathsf{A}}\|^{-1}$ *consists of not more than* $m-1$ *path-connected components. Then*

$$\mathcal{C}_{\mathsf{A}} \leq m - 1.$$

Proof. The proof is the same as in [dDS18b, Thm. 13]. □

In [dDS18b] we missed that we actually only need the assumptions in Theorem 12. We previously stated that A must be continuous, there is an $e \in \operatorname{lin} \mathsf{A}$ s.t. $e > 0$ on $\mathcal{X}$ and $\mathcal{X}$ has not more than $m-1$ components. This of course implies the assumptions in Theorem 12. The key step in the proof was that for any moment sequence s we find by Richter's Theorem (Theorem 6) a simplicial cone spanned by $s_{\mathsf{A}}(x_1), \dots, s_{\mathsf{A}}(x_m)$ containing s. Then two $s_{\mathsf{A}}(x_i)$ and $s_{\mathsf{A}}(x_j)$ lie in the same component of $\operatorname{range} s_{\mathsf{A}} \cdot \|s_{\mathsf{A}}\|^{-1}$ and can therefore be connected by a path. Following this path shrinks the simplicial cone until s is contained in its boundary, i.e., s needs only $m-1$ atoms.

3. The moment cones $\mathcal{S}_{\mathsf{A}}$ and $\mathcal{T}_{\mathsf{A}}$

In Definition 5 we defined the moment cones $\mathcal{S}_{\mathsf{A}}$ and $\mathcal{T}_{\mathsf{A}}$ and we already found $\mathcal{T}_{\mathsf{A}} \subseteq \mathcal{S}_{\mathsf{A}}$ and $\mathcal{S}_{\mathsf{A}}$ is a convex cone.

In the following we will "only" deal with moment sequences where we know that they have a representing mixture with finitely many components. That is the definition of $\mathcal{T}_{\mathsf{A}}$ in Definition 5. However, an application of the Richter Theorem (Theorem 6) shows that this is enough.

Definition 13. Set $\mathsf{B} := \{b_1, \dots, b_m\}$ where b_i is a function on $\mathcal{X} \times \Sigma$ defined by

$$b_i(x, \sigma) := \int_{\mathcal{X}} a_i(y) \; \mathrm{d}\delta_{x,\sigma} \qquad \forall (x,\sigma) \in \mathcal{X} \times \Sigma. \tag{14}$$

Example 14 (Gaussian distribution, Example 1 revisited)**.** From (5) and (7) we find for $a_i(x) = x^\alpha$ that $b_i(\xi, \sigma) := \int_{\mathbb{R}^n} x^\alpha \; \mathrm{d}\delta^G_{\xi,\sigma}(x)$ with $x^\alpha = x_1^{\alpha_1} \cdots x_n^{\alpha_n}$, $\alpha_i \in \mathbb{N}_0$, is a polynomial in $x_1, \dots, x_n$ and $\sigma_{i,j}$ of degree $|\alpha| = \alpha_1 + \dots + \alpha_n$.

Example 15 (Log-normal distribution, Example 2 revisited)**.** From (8) we find for $a_i(x) = x^i$ on $(0, \infty)$ that $b_i(\xi, \sigma) := \int_0^\infty x^i \; \mathrm{d}\delta^L_{\xi,\sigma}(x) = \xi^i \cdot e^{\frac{i^2\sigma^2}{2}}$.

B is well-defined by condition (b). Then any finite or infinite sums of components or continuous versions are measures on $\mathcal{X} \times \Sigma$. The Richter Theorem for mixtures of distributions reads as follows.

Theorem 16. *Let $\mathcal{X}$ and Σ be topological spaces, and let $\mathsf{B} = \{b_1, \dots, b_m\}$ be a finite set of functions on $\mathcal{X} \times \Sigma$, i.e., $\delta_{x,\sigma} \in \mathcal{M}_\mathsf{B}$ for all $(x, \sigma) \in \mathcal{X} \times \Sigma$. Then for every $\mu \in \mathcal{M}_\mathsf{B}$ there is a mixture with finitely many components $\mu' = (C, X, \bar{\sigma}) = \sum_{i=1}^k c_i \cdot \delta_{x_i,\sigma_i}$ with the same moment sequence and $k \leq m$, i.e.,*

$$\left(\int_{\mathcal{X}\times\Sigma} b_i(x,\sigma)\ \mathrm{d}\mu(x,\sigma)\right)_{i=1}^m \in \mathcal{T}_\mathsf{A} = \operatorname{range} T_{m,\mathsf{A}} = T_{m,\mathsf{A}}(\mathbb{R}^m_{\geq 0} \times \mathcal{X}^m \times \Sigma^m).$$

Proof. Apply the Richter Theorem (Theorem 6) to $\mathsf{B} = \{b_1, \dots, b_m\}$ on $\mathcal{X} \times \Sigma$. $\square$

So it is sufficient to deal "only" with moment sequences coming from finite mixtures.

Theorem 17. *Let $\mathsf{A} = \{a_1, \dots, a_m\}$ be linearly independent continuous functions on $\mathcal{X}$. Then*

i) *$\mathcal{T}_\mathsf{A}$ is a full-dimensional convex cone.*
ii) *$\operatorname{int} \mathcal{T}_\mathsf{A} = \operatorname{int} \mathcal{S}_\mathsf{A}$.*
iii) *Assume that*
   1) *$\mathcal{X}$ is a locally compact Hausdorff space,*
   2) *for every $x \in \mathcal{X}$ and $\sigma \in \Sigma$ there is a compact neighborhood $U_{x,\sigma} \subseteq \operatorname{supp} \delta_{x,\sigma}$ with $\delta_{x,\sigma}(U_{x,\sigma}) > 0$, and*
   3) *for every $f \in \operatorname{lin} \mathsf{A}$ with $f \geq 0$ on $\mathcal{X}$ and $f|_U = 0$ for a neighborhood U implies $f = 0$.*

   Then

$$\mathcal{T}_\mathsf{A} = \operatorname{int} \mathcal{S}_\mathsf{A} \cup \{0\}.$$

Proof. i): That $\mathcal{T}_\mathsf{A}$ is a cone is clear. That $\mathcal{T}_\mathsf{A}$ is convex follows from the Carathéodory Theorem for cones; see, e.g., [Roc72, Cor. 17.1.2]. To show that $\mathcal{T}_\mathsf{A}$ is full-dimensional, we take $x_1, \dots, x_m \in \mathcal{X}$ s.t. $s_\mathsf{A}(x_1), \dots, s_\mathsf{A}(x_m)$ are linearly independent (such x_i's exist since $\mathsf{A} = \{a_1, \dots, a_m\}$ is linearly independent). Let $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ s.t. $\sigma_i \to \sigma_0$ as $i \to \infty$. Then

$$\lim_{i\to\infty} \det(t_\mathsf{A}(x_1,\sigma_i), \dots, t_\mathsf{A}(x_m,\sigma_i)) = \det(s_\mathsf{A}(x_1), \dots, s_\mathsf{A}(x_m)) \neq 0$$

by condition (c) and continuity of the determinant, i.e., there is an $N \in \mathbb{N}$ s.t. $\det(t_\mathsf{A}(x_1,\sigma_N), \dots, t_\mathsf{A}(x_m,\sigma_N)) \neq 0$ and therefore $t_\mathsf{A}(x_1,\sigma_N), \dots, t_\mathsf{A}(x_m,\sigma_N)$ are linearly independent in $\mathbb{R}^m$ and $\mathcal{T}_\mathsf{A}$ is full-dimensional.

ii): From $\mathcal{T}_\mathsf{A} \subseteq \mathcal{S}_\mathsf{A}$ we get $\operatorname{int} \mathcal{T}_\mathsf{A} \subseteq \operatorname{int} \mathcal{S}_\mathsf{A}$. So we have to prove the reverse inclusion $\operatorname{int} \mathcal{T}_\mathsf{A} \supseteq \operatorname{int} \mathcal{S}_\mathsf{A}$. Let $s \in \operatorname{int} \mathcal{S}_\mathsf{A}$. Then there are $x_1, \dots, x_m \in \mathcal{X}$ and $c_i > 0$ s.t. $s = \sum_{i=1}^m c_i s_\mathsf{A}(x_i)$. So $s \in \operatorname{int} \operatorname{cone}(s_\mathsf{A}(x_1), \dots, s_\mathsf{A}(x_m))$. By (c) there exists $\sigma \in \Sigma$ s.t. $s \in \operatorname{int} \operatorname{cone}(t_\mathsf{A}(x_1,\sigma), \dots, t_\mathsf{A}(x_m,\sigma)) \subseteq \operatorname{int} \mathcal{T}_\mathsf{A}$.

iii): Since $0 \in \mathcal{T}_\mathsf{A}$ and $\operatorname{int} \mathcal{S}_\mathsf{A} = \operatorname{int} \mathcal{T}_\mathsf{A} \subseteq \mathcal{T}_\mathsf{A}$ by ii) we have $\operatorname{int} \mathcal{S}_\mathsf{A} \cup \{0\} \subseteq \mathcal{T}_\mathsf{A}$. So it is sufficient to prove the reverse inclusion $\mathcal{T}_\mathsf{A} \subseteq \operatorname{int} \mathcal{S}_\mathsf{A} \cup \{0\}$.

Assume this inclusion does not hold, i.e., $\mathcal{T}_\mathsf{A} \cap \partial^* \mathcal{S}_\mathsf{A} \neq \{0\}$ since $\mathcal{T}_\mathsf{A} \subseteq \mathcal{S}_\mathsf{A}$. Let $s \in \mathcal{T}_\mathsf{A} \cap \partial^* \mathcal{S}_\mathsf{A}$, $s \neq 0$; then $\mu = \sum_{i=1}^k c_i t_\mathsf{A}(x_i,\sigma_i)$ is a non-trivial representing measure of s (since $s \in \mathcal{T}_\mathsf{A}$) and there is an $f \geq 0$ in $\operatorname{lin} \mathsf{A} \setminus \{0\}$ s.t. $\int_\mathcal{X} f(x)\ \mathrm{d}\mu(x) = 0$ (since $s \in \partial^* \mathcal{S}_\mathsf{A}$; f is a separating hyperplane supporting $\mathcal{S}_\mathsf{A}$ at s). Let $U_{x_1,\sigma_1} \subseteq \operatorname{supp} \delta_{x_1,\sigma_1}$ be a compact neighborhood; then by continuity of f and 3) we have

$c := \max_{x\in U_{x_1,\sigma_1}} f(x) \in (0,\infty)$. Therefore, $U := U_{x_1,\sigma_1} \cap f^{-1}((c/2, 2c))$ is open in U_{x_1,σ_1} by continuity of f, i.e., $\delta_{x_1,\sigma_1}(U) > 0$. Then

$$0 = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x) \geq \int_{U_{x_1,\sigma_1}} f(x)\ \mathrm{d}\delta_{x_1,\sigma_1} \geq \frac{c}{2}\delta_{x_1,\sigma_1}(U) > 0.$$

This is a contradiction, i.e., $\mathcal{T}_{\mathsf{A}} \cap \partial^*\mathcal{S}_{\mathsf{A}} = \{0\}$ and therefore $\mathcal{T}_{\mathsf{A}} \subseteq \operatorname{int}\mathcal{S}_{\mathsf{A}} \cup \{0\}$. □

In iii) in the previous theorem we actually proved the following. It is a reformulation of Lemma 3 in [dDS18a].

Lemma 18. *Assume*

1) *$\mathcal{X}$ is a locally compact Hausdorff space,*
2) *for every $x \in \mathcal{X}$ and $\sigma \in \Sigma$ there is a compact neighborhood $U_{x,\sigma} \subseteq \operatorname{supp}\delta_{x,\sigma}$ with $\delta_{x,\sigma}(U_{x,\sigma}) > 0$, and*
3) *for every $f \in \operatorname{lin}\mathsf{A}$ with $f \geq 0$ on $\mathcal{X}$ and $f|_U = 0$ for a neighborhood U implies $f = 0$.*

Then

$$\mu \text{ is a representing measure of } s \in \mathcal{S}_{\mathsf{A}} \text{ with } \operatorname{int}\operatorname{supp}\mu \neq \emptyset \quad\Rightarrow\quad s \in \operatorname{int}\mathcal{S}_{\mathsf{A}}.$$

Example 19 (Gaussian mixtures, Example 1 revisited)**.** For the Gaussian mixtures we have $\mathcal{X} = \mathbb{R}^n$ (a locally compact Hausdorff space), $\operatorname{supp}\delta^G_{x,\sigma} = \mathcal{X} = \mathbb{R}^n$ for all $x \in \mathcal{X} = \mathbb{R}^n$ and $\sigma \in \Sigma \subseteq \mathbb{R}^{n\times n}$, the set of all symmetric non-singular matrices. Let A be a linearly independent set of holomorphic functions (e.g., poly-/monomials). Lemma 18 applies and every moment sequence s is an inner point of the moment cone $\mathcal{S}_{\mathsf{A}}$, i.e., the set of non-zero moment sequences from Gaussian mixtures is open.

Example 20 (Log-normal mixtures, Example 2 revisited)**.** For the Gaussian mixtures we have $\mathcal{X} = (0,\infty)^n$ (a locally compact Hausdorff space), $\operatorname{supp}\delta^L_{x,\sigma} = \mathcal{X} = (0,\infty)^n$ for all $x \in \mathcal{X} = (0,\infty)^n$ and $\sigma \in \Sigma \subseteq \mathbb{R}^{n\times n}$, the set of all symmetric non-singular matrices. Let A be a linearly independent set of holomorphic functions (e.g., poly-/monomials). Lemma 18 applies and every moment sequence s is an inner point of the moment cone $\mathcal{S}_{\mathsf{A}}$, i.e., the set of non-zero moment sequences from log-normal mixtures is open.

Of course, A being holomorphic can be weakened to condition 3) in Lemma 18.

In Theorem 17 ii) we actually showed that any $s \in \operatorname{int}\mathcal{T}_{\mathsf{A}}$ has a mixture representation with (at most) m components and all components have the same σ. In the following theorem we will show that we can represent large parts of $\mathcal{T}_{\mathsf{A}}$ by mixture representations with (at most) m components and all components have the same σ. σ must "just" be close enough to σ_0.

Definition 21. $\mathcal{T}_{\mathsf{A},\sigma} := T_{m,\mathsf{A}}(\mathbb{R}^m_{\geq 0} \times \mathcal{X}^m \times \{(\sigma,\dots,\sigma)\})$.

So $\mathcal{T}_{\mathsf{A},\sigma}$ is the (convex) set of all moment sequences s s.t. every s possesses a mixture representation $\sum_{i=1}^k c_i\delta_{x_i,\sigma}$ $(k \leq m)$ with at most m components and all components have the same σ.

Theorem 22. i) *$\mathcal{T}_{\mathsf{A},\sigma}$ is a convex cone for all $\sigma \in \Sigma$.*

ii) *Let $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ s.t. $\sigma_i \to \sigma_0$ as $i \to \infty$. Then*

$$\operatorname{int}\mathcal{T}_{\mathsf{A}} \cup \{0\} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{T}_{\mathsf{A},\sigma_i} \subseteq \bigcup_{\sigma\in\Sigma} \mathcal{T}_{\mathsf{A},\sigma}.$$

iii) *Let* $s_1, \dots, s_k \in \operatorname{int} \mathcal{T}_{\mathsf{A}}$ *be points and* $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ *s.t.* $\sigma_i \to \sigma_0$ *as* $i \to \infty$. *Then there exists an* $N \in \mathbb{N}$ *s.t.*

$$s_1, \dots, s_k \in \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i} \quad \forall i \geq N.$$

iv) *Let* $K \subset \operatorname{int} \mathcal{T}_{\mathsf{A}}$ *be compact and* $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ *s.t.* $\sigma_i \to \sigma_0$ *as* $i \to \infty$. *Then there exists an* $N \in \mathbb{N}$ *s.t.*

$$K \subset \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i} \quad \forall i \geq N.$$

v) *Let* $C \subset \operatorname{int} \mathcal{T}_{\mathsf{A}}$ *be a closed cone and* $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ *s.t.* $\sigma_i \to \sigma_0$ *as* $i \to \infty$. *Then there exists an* $N \in \mathbb{N}$ *s.t.*

$$C \subset \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i} \quad \forall i \geq N.$$

vi) *Assume conditions* 1), 2), *and* 3) *from Theorem* 17 iii) *hold and let* $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ *s.t.* $\sigma_i \to \sigma_0$ *as* $i \to \infty$. *Then*

$$\mathcal{T}_{\mathsf{A}} = \bigcup_{i\in\mathbb{N}} \mathcal{T}_{\mathsf{A},\sigma_i} = \bigcup_{\sigma\in\Sigma} \mathcal{T}_{\mathsf{A},\sigma}.$$

Proof. i): That $\mathcal{T}_{\mathsf{A}}$ is a cone is clear. That $\mathcal{T}_{\mathsf{A}}$ is convex follows from the Carathéodory Theorem for cones; see, e.g., [Roc72, Cor. 17.1.2].

ii): The proof follows the proof of Theorem 17 ii). Of course, $0 \in \mathcal{T}_{\mathsf{A},\sigma}$ for all $\sigma \in \Sigma$ and the second inclusion holds. So let $s \in \operatorname{int} \mathcal{T}_{\mathsf{A}}$. Then $s \in \operatorname{int} \mathcal{S}_{\mathsf{A}}$ by Theorem 17 ii) and there are $x_1, \dots, x_m \in \mathcal{X}$ and $c_i > 0$ s.t. $s = \sum_{i=1}^m c_i s_{\mathsf{A}}(x_i)$. So $s \in \operatorname{int} \operatorname{cone}(s_{\mathsf{A}}(x_1), \dots, s_{\mathsf{A}}(x_m))$. By (c) there exists $\sigma_i \in \Sigma$ s.t.

$$s \in \operatorname{int} \operatorname{cone}(t_{\mathsf{A}}(x_1, \sigma), \dots, t_{\mathsf{A}}(x_m, \sigma)) \subseteq \operatorname{int} \mathcal{T}_{\mathsf{A}}.$$

iii): As in ii) let $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ s.t. $\sigma_i \to \sigma_0$ as $i \to \infty$. By ii) for s_i there is an $N_i \in \mathbb{N}$ s.t. $s_i \in \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_l}$ for all $l \geq N_i$. Set $N := \max\{l_1, \dots, l_k\}$. Then $s_1, \dots, s_k \in \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N$.

iv): $\operatorname{conv} K$ is compact since K is compact and $\operatorname{conv} K \subset \operatorname{conv} \operatorname{int} \mathcal{T}_{\mathsf{A}} = \operatorname{int} \mathcal{T}_{\mathsf{A}}$ since $\operatorname{int} \mathcal{T}_{\mathsf{A}}$ is convex. Therefore, $\operatorname{dist}(\partial \mathcal{T}_{\mathsf{A}}, \operatorname{conv} K) > 0$ and there are $s_1, \dots, s_k \in \operatorname{int} \mathcal{T}_{\mathsf{A}}$ s.t. $\operatorname{conv} K \subseteq \operatorname{conv}\{s_1, \dots, s_k\}$. By iii) there is an $N \in \mathbb{N}$ s.t. $s_1, \dots, s_k \in \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N$. Since all $\operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i}$ are convex, we have $\operatorname{conv}\{s_1, \dots, s_k\} \subset \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N$. In conclusion we have

$$K \subseteq \operatorname{conv} K \subseteq \operatorname{conv}\{s_1, \dots, s_k\} \subset \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i} \quad \forall i \geq N.$$

v): Let $\mathbb{S}^m$ be the unit sphere in $\mathbb{R}^m$. Then $K = C \cap \mathbb{S}^m$ is closed and bounded (i.e., compact by the Heine–Borel Theorem) and generates C (i.e., $\operatorname{cone} K = C$). By iv) there is an $N \in \mathbb{N}$ s.t. $K \subset \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N$. Since $\mathcal{T}_{\mathsf{A},\sigma_i}$ are (convex) cones by i) we have the $C = \operatorname{cone} K \subset \operatorname{cone} \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i} = \operatorname{int} \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N$.

vi): From Theorem 17 ii) and iii) we have $\mathcal{T}_{\mathsf{A}} = \operatorname{int} \mathcal{S}_{\mathsf{A}} \cup \{0\} = \operatorname{int} \mathcal{T}_{\mathsf{A}} \cup \{0\}$. Then with ii) in this theorem we have

$$\mathcal{T}_{\mathsf{A}} \;=\; \operatorname{int} \mathcal{T}_{\mathsf{A}} \cup \{0\} \;\subseteq\; \bigcup_{i\in\mathbb{N}} \mathcal{T}_{\mathsf{A},\sigma_i} \;\subseteq\; \bigcup_{\sigma\in\Sigma} \mathcal{T}_{\mathsf{A},\sigma} \;\subseteq\; \mathcal{T}_{\mathsf{A}}. \qquad \square$$

4. Set of atoms and identifiability/uniqueness/determinacy

We have seen that any $s \in \mathcal{T}_{\mathsf{A}}$ has a finite mixture representation (by definition) and we want to know the possible positions (x, σ) s.t. $\delta_{x,\sigma}$ appears in any such representation.

Definition 23. Let $s \in \mathcal{T}_\mathsf{A}$. The set of atoms (components) $\mathcal{W}(s)$ is defined by

$$\mathcal{W}(s) := \left\{ (x,\sigma) \;\middle|\; \exists c_i, c > 0, (x_i,\sigma_i) \in \mathcal{X} \times \Sigma : s = c \cdot t_\mathsf{A}(x,\sigma) + \sum_{i=1}^k c_i t_\mathsf{A}(x_i,\sigma_i) \right\}.$$

So $\mathcal{W}(s)$ is the set of all $(x,\sigma) \in \mathcal{X} \times \Sigma$ s.t. $\delta_{x,\sigma}$ appears in a mixture representation of s.

Definition 24. Let $s \in \mathcal{T}_\mathsf{A}$. s is called *determined* if it has only one mixture representation. Otherwise s is called *indeterminate.*

The following theorem summarizes properties of $\mathcal{W}(s)$ and determinacy. It is a reformulation of Theorems 16 and 19 in [dDS18a].

Theorem 25. *Let $s \in \mathcal{T}_\mathsf{A}$.*

i) $s \in \operatorname{int} \mathcal{T}_\mathsf{A} \;\Leftrightarrow\; \mathcal{W}(s) = \mathcal{X} \times \Sigma$.

ii) *s is indeterminate* $\;\Leftrightarrow\;$ $\{t_\mathsf{A}(x,\sigma) \,|\, (x,\sigma) \in \mathcal{W}(s)\}$ *is linearly dependent.*

iii) *Assume that*

  1) *$\mathcal{X}$ is a locally compact Hausdorff space,*
  2) *for every $x \in \mathcal{X}$ and $\sigma \in \Sigma$ there is a compact neighborhood $U_{x,\sigma} \subseteq \operatorname{supp} \delta_{x,\sigma}$ with $\delta_{x,\sigma}(U_{x,\sigma}) > 0$, and*
  3) *for every $f \in \operatorname{lin} \mathsf{A}$ with $f \geq 0$ on $\mathcal{X}$ and $f|_U = 0$ for a neighborhood U implies $f = 0$.*

Then every $s \in \mathcal{T}_\mathsf{A} \setminus \{0\}$ is indeterminate and $\mathcal{W}(s) = \mathcal{X} \times \Sigma$ for all $s \in \mathcal{T}_\mathsf{A} \setminus \{0\}$.

Proof. i) "$\Rightarrow$": Let $(x,\sigma) \in \mathcal{X} \times \Sigma$. Since $s \in \operatorname{int} \mathcal{T}_\mathsf{A}$ there is an $\varepsilon > 0$ s.t. $s' := s - \varepsilon \cdot t_\mathsf{A}(x,\sigma) \in \operatorname{int} \mathcal{T}_\mathsf{A}$. Then s' has a finite mixture representation μ' and $\mu := \mu' + \varepsilon \cdot \delta_{x,\sigma}$ is a mixture presentation of s containing $\delta_{x,\sigma}$.

i) "$\Leftarrow$": Let $(x_i,\sigma_i) \in \mathcal{W}(s) = \mathcal{X} \times \Sigma$ $(i = 1, \dots, m)$ s.t. $(t_\mathsf{A}(x_1,\sigma_1), \dots, t_\mathsf{A}(x_m,\sigma_m))$ has full rank. Let μ_i be representing mixtures of s s.t. every μ_i contains the component δ_{x_i,σ_i}. Then

$$\mu := \frac{1}{m} \sum_{i=1}^m \mu_i = \sum_{i=1}^m c_i \delta_{x_i,\sigma_i} + \sum_{j=1}^k d_j \delta_{y_j,\sigma_j'}$$

is a representing mixture of s which contains all δ_{x_i,σ_i}. Then the map

$$S(\gamma_1, \dots, \gamma_m) := \sum_{i=1}^m \gamma_i t_\mathsf{A}(x_i,\sigma_i) + \sum_{j=1}^k d_j t_\mathsf{A}(y_j,\sigma_j')$$

maps a neighborhood $B_\varepsilon((c_1, \dots, c_m)) \subset (0,\infty)^n$ with $0 < \varepsilon < \min\{c_1, \dots, c_m\}$ to a neighborhood of s since the $t_\mathsf{A}(x_i,\sigma_i)$ are linearly independent, i.e., $s \in \operatorname{int} \mathcal{T}_\mathsf{A}$.

ii): Apply (i) $\Leftrightarrow$ (ii) in [dDS18a, Thm. 19].

iii): By Theorem 17 iii) $\mathcal{T}_\mathsf{A} \setminus \{0\} = \operatorname{int} \mathcal{S}_\mathsf{A}$ is open and i) and ii) apply. □

Example 26 (Examples 1 and 2 revisited)**.** For the Gaussian and log-normal distributions the conditions 1), 2), and 3) are fulfilled, i.e., every moment sequence is indeterminate and any component $\delta^G_{x,\sigma}$ or $\delta^L_{x,\sigma}$, respectively, can appear in a representing mixture.

That for any $s \in \operatorname{int} \mathcal{T}_\mathsf{A}$ any $\delta_{x,\sigma}$ appears in a representing mixture is only possible since the number of components is not restricted. So we need to learn more about the number of components.

5. The Carathéodory number $\mathcal{C}_{\mathsf{A}}^M$ for mixtures of distributions

Since (by definition or Theorem 16) every $s \in \mathcal{T}_{\mathsf{A}}$ has a mixture representation we can define the Carathéodory number of mixtures similar to Definition 7.

Definition 27. Let $s \in \mathcal{T}_{\mathsf{A}}$. Define the *Carathéodory number $\mathcal{C}_{\mathsf{A}}^M(s)$ of mixtures of* s by

$$\mathcal{C}_{\mathsf{A}}^M(s) \ := \ \min \{k \in \mathbb{N} \,|\, s \text{ has a mixture representation with } k \text{ components}\}.$$

The *Carathéodory number $\mathcal{C}_{\mathsf{A}}^M$ of mixtures* is defined by

$$\mathcal{C}_{\mathsf{A}}^M \ := \ \max_{s\in\mathcal{T}_{\mathsf{A}}} \ \mathcal{C}_{\mathsf{A}}^M(s).$$

$\mathcal{C}_{\mathsf{A}}^M(s)$ and $\mathcal{C}_{\mathsf{A}}^M$ are well-defined by Theorem 6 or equivalently Theorem 16 since $0 \leq \mathcal{C}_{\mathsf{A}}^M(s) \leq \mathcal{C}_{\mathsf{A}}^M \leq m$ and $\mathcal{C}_{\mathsf{A}}(s), \mathcal{C}_{\mathsf{A}} \in \mathbb{N}_0$. We will shortly see in Example 31 that not necessarily $\mathcal{C}_{\mathsf{A}}^M \leq \mathcal{C}_{\mathsf{A}}$ even though $\mathcal{T}_{\mathsf{A}} \subseteq \mathcal{S}_{\mathsf{A}} = \operatorname{range} S_{\mathcal{C}_{\mathsf{A}},\mathsf{A}}$ holds.

In important cases, e.g., Gaussian and log-normal mixtures (Examples 1 and 2), the moment cone has no boundary points despite 0. So "standard" methods to bound $\mathcal{C}_{\mathsf{A}}^M$ in [dDS18b] and [RS18] cannot be applied. These "standard" methods are, e.g., "taking an inner point, removing an atom to get to the boundary and describe the boundary" or "close the moment cone by going from $\mathbb{R}^n$ to projective space $\mathbb{P}^n$ and to homogeneous polynomials". In all these cases, a boundary point $s \neq 0$ would imply that there is an $f \in \operatorname{lin} \mathsf{A}$, $f \geq 0$, such that $\operatorname{supp} \mu \subseteq \mathcal{W}(s) \subseteq \mathcal{Z}(f) := \{x \in \mathcal{X} \,|\, f(x) = 0\} \neq \mathcal{X}$ for all representing measures μ of s. But this is not possible as long as conditions 1), 2), and 3) shall hold and $\operatorname{int} \operatorname{supp} \delta_{x,\sigma} \neq \emptyset$; see Theorem 25 iii). So recent methods in [dDS18b] and [RS18] do not apply. Theorem 22 fills the gap. But let us start with the lower bounds on $\mathcal{C}_{\mathsf{A}}^M$.

Definition 28. Let $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ and $\Sigma \subseteq \mathbb{R}^{n_2}$ be open. Furthermore, let b_i from Definition 13 be C^1-functions. We define

$$\mathcal{N}_{\mathsf{A}}^M := \min\{k \in \mathbb{N} \,|\, DT_{k,\mathsf{A}} \text{ has full rank}\}.$$

$DT_{k,\mathsf{A}}$ is the total derivative of $T_{k,\mathsf{A}}$.

Example 29. For the Gaussian distribution we have $\mathcal{X} = \mathbb{R}^n$ and for the log-normal distribution we have $\mathcal{X} = (0,\infty)^n$; see Examples 1 and 2. In both cases Σ is the set of all symmetric non-singular matrices in $\mathbb{R}^{n\times n}$, e.g., Σ is an open $\frac{n(n+1)}{2}$-dimensional smooth manifold, i.e., Remark 11 applies.

We have the following lower bound on $\mathcal{C}_{\mathsf{A}}^M$.

Theorem 30. *Let $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ and $\Sigma \subseteq \mathbb{R}^{n_2}$ be open. Furthermore, let b_i from Definition 13 be C^r-functions with $r > \mathcal{N}_{\mathsf{A}}(n_1+n_2+1) - m$. Then*

$$\left\lceil \frac{m}{n_1+n_2+1} \right\rceil \leq \ \mathcal{N}_{\mathsf{A}}^M \ \leq \ \mathcal{C}_{\mathsf{A}}^M.$$

Proof. Apply Theorem 10 [dDS18b, Thm. 27] with $\mathcal{X} \times \Sigma \subseteq \mathbb{R}^{n_1+n_2}$. □

See also Remark 11 for extensions of $\mathcal{X}$ and Σ to differentiable manifolds. The previous theorem suggests that there are cases where $\mathcal{C}_{\mathsf{A}}^M \not\leq \mathcal{C}_{\mathsf{A}}$.

Example 31 ($\mathcal{C}_{\mathsf{A}}^M \not\leq \mathcal{C}_{\mathsf{A}}$; see [dDS18b, Exm. 16 and Rem. 17])**.** Let $\varphi = (\varphi_i)_{i=1}^m$ be the coordinate functions of a space filling curve [Sag94, Ch. 5 and 7], i.e., $\varphi : [0,1] \to [0,1]^m$ are continuous. Extend all φ_i continuously to $\mathbb{R}$ s.t. $\operatorname{supp} \varphi_i \subseteq [-1,2]$. For the Gaussian distributions (Example 1) we can then interchange differentiation and integration in (14) in Definition 13 by applying a result of Lebesgue (see, e.g., [Gru09, Lem. 2.8]) and we get that all b_i's are C^∞. Therefore Theorem 30 holds and we get with Example 29 that $n+1+\frac{n(n+1)}{2} = \frac{(n+2)(n+1)}{2}$ and $\left\lceil \frac{2m}{(n+2)(n+1)} \right\rceil \leq \mathcal{C}_{\mathsf{A}}^M$, i.e., for $2m > (n+2)(n+1)$ we have $\mathcal{C}_{\mathsf{A}} = 1 < \mathcal{C}_{\mathsf{A}}^M$.

If we are interested in representations with fixed σ for all distributions, we need at least $\left\lceil \frac{m}{n_1+1} \right\rceil$ distributions $\delta_{x_i,\sigma}$. And when we want a presentation s.t. all $\sigma_1 = \cdots = \sigma_k = \sigma$ are the same but we are allowed to choose σ freely, then we need at least $\left\lceil \frac{m-n_2}{n_1+1} \right\rceil$ distributions. Apply Theorem 10 or modify the proof in [dDS18b, Thm. 27] to prove these.

Let us now treat the upper bound estimates. We already established $\mathcal{C}_{\mathsf{A}}^M \leq m$ in Theorem 16. We can tighten this.

Theorem 32. *Let* A, $\mathcal{X}$, Σ, $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$, *and* $\delta_{x,\sigma}$ *s.t.* $\sigma_i \to \sigma_0$, *there exists an* $e \in \operatorname{lin}\{b_1,\dots,b_m\}$ *from Definition* 13 *and* $N \in \mathbb{N}$ *with* $e(x,\sigma_i) > 0$ *for all* $x \in \mathcal{X}$ *and* $i \geq N$, *and* $\operatorname{range} t_{\mathsf{A}} \cdot \|t_{\mathsf{A}}\|^{-1}$ *consists of not more than* $m-1$ *path-connected components. Then*

$$\mathcal{C}_{\mathsf{A}}^M \leq m-1.$$

Proof. Let $s \in \mathcal{T}_{\mathsf{A}}$. Then by Theorem 22 ii) there is an $N' \in \mathbb{N}$ s.t. $s \in \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq N'$. Apply Theorem 12 to $\mathcal{T}_{\mathsf{A},\sigma_i}$ for some $i \geq \max\{N, N'\}$. □

Theorem 22 can be used to bound $\mathcal{C}_{\mathsf{A}}^M$.

Theorem 33. *Let* $(\sigma_i)_{i\in\mathbb{N}} \subseteq \Sigma$ *s.t.* $\sigma_i \to \sigma_0$ *as* $i \to \infty$ *and there exist* $C, N \in \mathbb{N}$ *s.t.* $\mathcal{C}_{\{b_1(\,\cdot\,,\sigma_i),\dots,b_m(\,\cdot\,,\sigma_i)\}} \leq C$ *for all* $i \geq N$. *Then*

$$\mathcal{C}_{\mathsf{A}}^M \leq C.$$

Proof. Let $s \in \mathcal{T}_{\mathsf{A}}$. Then by Theorem 22 ii) there is an $N' \in \mathbb{N}$ s.t. $s \in \mathcal{T}_{\mathsf{A},\sigma_i}$ for all $i \geq \max\{N, N'\}$, i.e., $\mathcal{C}_{\mathsf{A}}^M(s) \leq C$ since $\mathcal{C}_{\{b_1(\,\cdot\,,\sigma_i),\dots,b_m(\,\cdot\,,\sigma_i)\}} \leq C$ for all $i \geq N$. Since s was arbitrary, we have $\mathcal{C}_{\mathsf{A}}^M \leq C$. □

Let us give an application to the most common cases: Gaussian and log-normal distributions (Examples 1 and 2). Let us start with the following remark.

Remark 34. Let $\mathsf{A}_{n,d} := \{x^\alpha \mid \alpha \in \mathbb{N}_0^n \wedge |\alpha| \leq d\}$ be the monomials of degree at most d in n variables and $(\sigma_i)_{i\in\mathbb{N}} \subset \Sigma$ with $\sigma_i := i^{-1}\mathrm{id}$, id the identity matrix. For the Gaussian distributions $\delta_{x,\sigma}^G$ we find from (7) that $b_\alpha(x, i^{-1}\mathrm{id}) := \int_{\mathbb{R}^n} y^\alpha \,\mathrm{d}\delta_{x,i^{-1}\mathrm{id}}^G(y)$ is a polynomial in $x_1,\dots,x_n$ with leading term x^α. So

$$\operatorname{lin}\{b_\alpha(x, i^{-1}\mathrm{id}) \mid \alpha \in \mathbb{N}_0^n \wedge |\alpha| \leq d\} = \operatorname{lin} \mathsf{A}_{n,d}. \tag{15}$$

Since the Carathéodory number does not depend on the choice of basis functions spanning $\operatorname{lin}\{b_\alpha(x, i^{-1}\mathrm{id})\}$ we can apply Theorem 33 with results from previous studies of Carathéodory numbers from Dirac measures; see, e.g., [dDS18b] and [RS18].

For the log-normal distribution $\delta^L_{x,\sigma}$ we find the same: (15) holds by (8). But we have $\mathcal{X} = (0,\infty)$.

Let us apply the previous remark.

Theorem 35. *Let $\mathsf{A}_{n,d} = \{x^\alpha \,|\, \alpha \in \mathbb{N}_0^n \wedge |\alpha| \leq d\}$ be the monomials of degree at most d in n variables. Then for Gaussian and log-normal mixtures we have*

$$\mathcal{C}^M_{\mathsf{A}_{n,d}} \leq \mathcal{C}_{\mathsf{A}_{n,d}}.$$

Proof. Follows from (15) and Theorem 33. □

Let us give some explicit applications of the previous theorem. We have $\mathcal{C}^M_{\mathsf{A}_{2,2d-1}} \leq \frac{3}{2}d(d-1)+1$ by [RS18] and for the one-dimensional Gaussian mixture we have the following.

Corollary 36. *Let $\mathsf{A} = \{1, x, \dots, x^d\}$ on $\mathbb{R}$, $d \in \mathbb{N}$. For the Gaussian mixtures we have*

$$\left\lceil \frac{d+1}{3} \right\rceil \quad \leq \quad \mathcal{C}^M_{\mathsf{A}} \quad \leq \quad \left\lceil \frac{d+1}{2} \right\rceil,$$

and every moment sequence s coming from a linear combination of Gaussian measures can be written as

$$s = \sum_{i=1}^k c_i \cdot s_{\mathsf{A},\sigma}(x_i) \quad \text{with} \quad k \leq \left\lceil \frac{d+1}{2} \right\rceil \quad \text{and} \quad \text{some} \quad \sigma = \sigma(s) > 0.$$

Equivalently, every moment sequence s from a Gaussian mixture has a Gaussian mixture representation

$$F(x) = \sum_{i=1}^k c_i \cdot e^{-\frac{(x-x_i)^2}{2\sigma^2}} \quad \text{with} \quad k \leq \left\lceil \frac{d+1}{2} \right\rceil \quad \text{and} \quad \text{some} \quad \sigma = \sigma(s) > 0.$$

Proof. $\mathcal{C}^M_{\mathsf{A}} \geq \left\lceil \frac{d+1}{3} \right\rceil$ follows from Theorem 30 with $n_1 = n_2 = 1$ and the upper bound follows from Theorem 35 with Theorem 8. □

For the one-dimensional log-normal distribution we will even have a more general result since it only lives on $(0,\infty)$; see Theorem 41.

For systems $\mathsf{A} \subset \mathbb{R}[x_1,\dots,x_n]$ with gaps, the application of previous results is more involved. (15) no longer holds. E.g., for $\mathsf{A} = \{1, x^2, x^3, x^5, x^6\}$ on $\mathbb{R}$ we get $b_0(x,\sigma) = 1$, $b_2(x,\sigma) = x^2+\sigma^2$, $b_3(x,\sigma) = x^3+3\sigma^2 x$, $b_5(x,\sigma) = x^5+10\sigma^2x^3+15\sigma^4 x$, $b_6(x,\sigma) = x^6 + 15\sigma^2 x^4 + 45\sigma^4 x^2 + 15\sigma^6$, so

$$\operatorname{lin}\{b_i(x,\sigma)\}_{i=1}^6 = \operatorname{lin}\{1,\ \ x^2,\ \ x^3+3\sigma^2 x,\ \ x^5 - 15\sigma^4 x,\ \ x^6 + 15\sigma^2 x^4\},$$

i.e., we always have contributions from x and x^4.

Systems with gaps, especially the univariate case, were treated in [dDS18b]. For $\mathsf{A} = \{1, x^2, x^3, x^5, x^6\}$ on $\mathbb{R}$ we found that $\mathcal{C}_{\mathsf{A}} = 3$ [dDS18b, Exm. 46]. Theorem 16 gives $\mathcal{C}^M_{\mathsf{A}} \leq 5$ while Theorem 12 gives a bound of $\mathcal{C}^M_{\mathsf{A}} \leq 4$. We will show with the following results, at least $\mathcal{C}^M_{\mathsf{A}}(s) \leq 3$ for almost every $s \in \mathcal{T}_{\mathsf{A}}$; see Example 39. At first we will show that a k-atomic Dirac measure (C,X) s.t. $DS_{k,\mathsf{A}}(C,X)$ has full rank gives a mixture with at most k components.

Theorem 37. *Let $\mathsf{A} \in C^1$ s.t. $b_i(x,\sigma\mathrm{id})$ and $\partial_j b_i(x,\sigma\mathrm{id})$ are continuous in $\sigma \in [0,\infty)$ and $x \in \mathbb{R}^n$ for all $i = 1,\dots,m$ and $j = 1,\dots,n$, and let $s \in \mathcal{S}_{\mathsf{A}}$ s.t. s has a k-atomic representing measure (C_s, X_s) with $DS_{k,\mathsf{A}}(C_s,X_s)$ has full rank. Then*

$$\mathcal{C}^M_{\mathsf{A}}(s) \leq k.$$

Proof. Since s has a k-atomic representing measure (C_s, X_s) s.t. $DS_{k,\mathsf{A}}(C_s, X_s)$ has full rank, $s \in \operatorname{int} \mathcal{S}_\mathsf{A} \subset \mathcal{T}_\mathsf{A}$ by Theorem 17. Since $DS_{k,\mathsf{A}}(C_s, X_s)$ has full rank, pick m variables $y = (y_1, \dots, y_m)$ from $c_1, \dots, c_k$ and $x_{1,1}, \dots, x_{k,n}$ s.t. $D_y S_{k,\mathsf{A}}(C_s, X_s) \in \mathbb{R}^{m\times m}$ is non-singular. Since $b_i(x, \sigma \mathrm{id})$ and $\partial_j b_i(x, \sigma\mathrm{id})$ are continuous in $\sigma \in [0, \infty)$ and $x \in \mathbb{R}^n$ for all $i = 1, \dots, m$ and $j = 1, \dots, n$, $D_y T_{k,\mathsf{A}}(C, X, \sigma(\mathrm{id}, \dots, \mathrm{id})) = D_y S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C, X)$ is continuous in $\sigma \in [0, \infty)$, $c_i \in [0, \infty)$, and $x \in \mathbb{R}^n$. Therefore, there is an $\varepsilon > 0$ s.t. $D_y S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}$ is non-singular for all $(\sigma, C, X) \in K_\varepsilon := [0, \varepsilon] \times \overline{B_\varepsilon(C_s, X_s)}$ since the determinant is continuous in the entries of the matrix. Denote by $\tau_1(\sigma, C, X) \le \cdots \le \tau_n(\sigma, C, X)$ the singular values of $D_y S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C, X)$. Since the singular values also depend continuously on the matrix, they depend continuously on $(\sigma, C, X) \in K_\varepsilon$. Since $\tau_i(\sigma, C, X)$ are continuous, they are bounded from above on the compact set K_ε. But since $\det(D_y S_{k,\mathsf{A}_{\sigma\mathrm{id}}}(C, X)) = \pm\tau_1(\sigma, C, X) \cdots \tau_n(\sigma, C, X) \neq 0$, τ_1 is non-zero on K_ε and there is a $\tau_{\min} > 0$ s.t.

$$\inf_{(\sigma,C,X)\in K_\varepsilon} \tau_1(\sigma, C, X) = \min_{(\sigma,C,X)\in K_\varepsilon} \tau_1(\sigma, C, X) \ge \tau_{\min}.$$

But this means that $B_r(S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C_s, X_s)) \subseteq S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(B_\varepsilon(C_s, X_s))$ for all $r \in (0, \min\{\varepsilon, \tau_{\min}\varepsilon\}) \;\wedge\; \sigma \in [0, \varepsilon]$. Fix $r \in (0, \min\{\varepsilon, \tau_{\min}\varepsilon\})$. Then $s \in B_r(S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C_s, X_s)) \subseteq S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(B_\varepsilon(C_s, X_s))$ for all $\sigma \in (0, \varepsilon)$ s.t. $\|s - S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C_s, X_s)\| < r$, i.e., $s = S_{k,\{b_i(\,\cdot\,,\sigma\mathrm{id})\}_i}(C, X) = T_{k,\mathsf{A}}(C, X, (\sigma\mathrm{id}, \dots, \sigma\mathrm{id}))$ for a $(C, X) \in B_\varepsilon(C_s, X_s)$. □

Note that in the proof of Theorem 37 the use of the multiple of id is arbitrary. Any non-singular symmetric matrix will do, just insert a basis transformation on $\mathbb{R}^n$. From Theorem 37 we get the following.

Theorem 38. *Let* $\mathsf{A} = \{a_1, \dots, a_m\}$ *in* $C^r(\mathbb{R}^n, \mathbb{R})$ *with* $r > \mathcal{N}_\mathsf{A} \cdot (n + 1) - m$ *s.t.* $b_i(x, \sigma\mathrm{id})$ *and* $\partial_j b_i(x, \sigma\mathrm{id})$ *are continuous in* $\sigma \in [0, \infty)$ *and* $x \in \mathbb{R}^n$ *for all* $i = 1, \dots, m$ *and* $j = 1, \dots, n$. *Then*

$$\mathcal{C}^M_\mathsf{A}(s) \quad\le\quad \mathcal{C}_\mathsf{A}(s) \quad\le\quad \mathcal{C}_\mathsf{A} \qquad \forall s \in \mathcal{T}_\mathsf{A}\ \lambda^n\text{-a.e.} \tag{16}$$

and the interior of the set where (16) *holds is dense in* $\mathcal{T}_\mathsf{A}$.

Proof. From Sard's Theorem [Sar42] we know that the set of singular values is of n-dimensional Lebesgue measure zero and Theorem 37 applies to the regular values, i.e., moment sequences s.t. all representing measures (C, X) have full rank $DS_{k,\mathsf{A}}(C, X)$. □

The open problem is: Can we ensure that any moment sequence has a representing measure (C, X) with full rank $DS_{k,\mathsf{A}}(C, X)$ with at most $\mathcal{C}_\mathsf{A}$ atoms? If we allow more atoms, this is true by [dDS18a, Lem. 36]. But this raises the Carathéodory bound. Let us give an example of Theorem 38.

Example 39. Let $\mathsf{A} = \{1, x^2, x^3, x^5, x^6\}$ on $\mathbb{R}$; see [dDS18b, Exm. 46]. There we found that $\mathcal{C}_\mathsf{A} = 3$. It is easily seen that A fulfills all condition in Theorem 38 (resp., Theorem 37) and therefore $\mathcal{C}^M_\mathsf{A}(s) \le 3$ for $s \in \mathcal{T}_\mathsf{A}$ λ^n-a.e. and the lower bound $\lceil \frac{5}{3} \rceil = 2 \le \mathcal{C}^M_\mathsf{A}$ holds because of Theorem 30.

We end this study with the general one-dimensional result applied to log-normal mixtures. It uses the fact that $x \in (0, \infty)$ and therefore a prior one-dimensional result [dDS18b, Lem. 40] can be applied.

Theorem 40. *Let* $m \in \mathbb{N}$ *and* $d_1, \dots, d_m \in \mathbb{N}_0$ *be such that* $d_1 < \dots < d_m$ *and* $\mathsf{A} = \{x^{d_1}, \dots, x^{d_m}\}$ *on* $\mathcal{X} = (0, \infty)$. *Then* $\mathcal{C}_\mathsf{A} = \lceil \frac{m}{2} \rceil$.

Proof. That $\mathcal{C}_\mathsf{A} \geq \lceil \frac{m}{2} \rceil$ follows from Theorem 10.

For $\mathcal{C}_\mathsf{A} \leq \lceil \frac{m}{2} \rceil$ let $s \in \mathcal{S}_\mathsf{A}$. By Richter's Theorem (Theorem 6) there is a k-atomic representing measure ($k \leq m$): $s = \sum_{i=1}^k c_i \cdot s_\mathsf{A}(x_i) = \sum_{i=1}^k \frac{c_i}{x_i} s_{x\mathsf{A}}(x_i)$ with $x_i \in (0, \infty)$, where $x\mathsf{A} = \{x^{d_1+1}, \dots, x^{d_m+1}\}$. Hence, w.l.o.g. we can assume $d_1 > 0$.

Let $d_1 > 0$ and we treat the extended homogeneous system

$$\mathsf{B} = \{y^{d_{m+1}}, x^{d_1} y^{d_{m+1}-d_1}, \dots, x^{d_{m+1}}\}$$

with $d_{m+1} = 2d_m$ on $\overline{\mathcal{X}}$, i.e., $(x, y) = (0, 1)$ is $x = 0$ on $\mathbb{R}$ and $(x, y) = (1, 0)$ is ∞ on $\mathbb{R}$. Then $\mathcal{S}_\mathsf{B}$ on $\overline{\mathcal{X}}$ is closed and pointed by [dDS18b, Prop. 8]. Set $\overline{s} = (s_i)_{i=0}^{m+1} = \sum_{i=1}^k c_i \cdot s_\mathsf{B}((x_i, 1))$; we added the moments s_0 and s_{m+1} to $s = (s_i)_{i=1}^m$. Again, by [dDS18b, Prop. 8] we have that

$$\begin{aligned} \overline{s}' := \overline{s} &- s_\mathsf{B}(1,0) \cdot \sup\{c \in \mathbb{R} \,|\, (\overline{s} - c \cdot s_\mathsf{B}(0,1)) \in \mathcal{S}_\mathsf{B}\} \\ &- s_\mathsf{B}(0,1) \cdot \sup\{c \in \mathbb{R} \,|\, (\overline{s} - c \cdot s_\mathsf{B}(1,0)) \in \mathcal{S}_\mathsf{B}\} \in \partial \mathcal{S}_\mathsf{B} \end{aligned}$$

is a boundary moment sequence of $\mathcal{S}_\mathsf{B}$ and by construction of $\overline{s}'$ every representing measure of $\overline{s}'$ does neither contain $(1, 0)$ nor $(0, 1)$: $\overline{s}' = \sum_{i=1}^k c_i' \cdot s_\mathsf{B}(x_i', 1)$ with $0 < x_1' < \dots < x_k' < \infty$. Since $\overline{s}'$ is a boundary point, $DS_{k,\mathsf{B}}((c_1', \dots, c_k'), (x_1', \dots, x_k'))$ is singular and from [dDS18b, Lem. 40/43] we get $k \leq \lceil \frac{m}{2} \rceil < \lceil \frac{m+2}{2} \rceil$. But since $s_\mathsf{B}(0, 1) = (1, 0, \dots, 0)$ and $s_\mathsf{B}(1, 0) = (0, \dots, 0, 1)$ the s_i for $i = 1, \dots, m$ in $\overline{s}$, $\overline{s}'$, and s are not altered, i.e., s has the k-atomic representing measure $((c_1', \dots, c_k'), (x_1', \dots, x_m'))$ with $0 < x_1' < \dots < x_k' < \infty$ and $k \leq \lceil \frac{m}{2} \rceil$. □

Theorem 41. *Let* $m \in \mathbb{N}$ *and* $d_1, \dots, d_m \in \mathbb{N}_0$ *be such that* $d_1 < \dots < d_m$ *and* $\mathsf{A} = \{x^{d_1}, \dots, x^{d_m}\} \subset \mathbb{R}[x]$. *Then for the log-normal distribution we have*

$$\left\lceil \frac{m}{3} \right\rceil \quad \leq \quad \mathcal{C}_\mathsf{A}^M \quad \leq \quad \left\lceil \frac{m}{2} \right\rceil.$$

Proof. The lower bound follows from Theorem 30. For the upper bound let $s \in \mathcal{T}_\mathsf{A}$. By Theorem 22(vi) there is a $\sigma > 0$ such that $s \in \mathcal{T}_{\mathsf{A},\sigma}$, i.e., we are in the one-dimensional setup of Theorem 40 which gives $\mathcal{C}_\mathsf{A}^M(s) \leq \lceil \frac{m}{2} \rceil$. □

Example 42. Let $\mathsf{A} = \{1, x, x^2, x^{17}, x^{1863}, x^{25\,376}\}$. Then by using Theorem 41 we find that every moment sequence from a log-normal mixture has another log-normal mixture representation with at most three components.

Acknowledgment

The author is grateful to Professor K. Schmüdgen for valuable discussions on the subject of this paper.

References

[ABB+17] C. Améndola, N. Bliss, I. Burke, C. R. Gibbons, M. Helmer, S. Hoşten, E. D. Nash, J. I. Rodriguez, and D. Simolkin, *The maximum likelihood degree of toric varieties*, arXiv:1703.02251v2, (2017).

[AFS16] Carlos Améndola, Jean-Charles Faugère, and Bernd Sturmfels, *Moment varieties of Gaussian mixtures*, J. Algebr. Stat. **7** (2016), no. 1, 14–28, DOI 10.18409/jas.v7i1.42. MR3529332

[AK62] N. I. Aheizer and M. Krein, *Some questions in the theory of moments*, translated by W. Fleming and D. Prill. Translations of Mathematical Monographs, Vol. 2, American Mathematical Society, Providence, R.I., 1962. MR0167806

[Akh65] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*, Translated by N. Kemmer, Hafner Publishing Co., New York, 1965. MR0184042

[ARS17] C. Améndola, K. Ranestad, and B. Sturmfels, *Algebraic identifiability of gaussian mixtures*, arXiv:1612.01129v2, (2017).

[BS73] Fischer Black and Myron Scholes, *The pricing of options and corporate liabilities*, J. Polit. Econ. **81** (1973), no. 3, 637–654, DOI 10.1086/260062. MR3363443

[CF96a] Raúl E. Curto and Lawrence A. Fialkow, *Flat extensions of positive moment matrices: recursively generated relations*, Mem. Amer. Math. Soc. **136** (1998), no. 648, x+56, DOI 10.1090/memo/0648. MR1445490

[CF96b] Raúl E. Curto and Lawrence A. Fialkow, *Solution of the truncated moment problem for flat data*, Mem. Amer. Math. Soc., vol. 119, American Mathematical Society, Providence, Rhode Island, 1996.

[CF00] Raúl E. Curto and Lawrence A. Fialkow, *The truncated complex K-moment problem*, Trans. Amer. Math. Soc. **352** (2000), no. 6, 2825–2855, DOI 10.1090/S0002-9947-00-02472-7. MR1661305

[CF05] Raúl E. Curto and Lawrence A. Fialkow, *Truncated K-moment problems in several variables*, J. Operator Theory **54** (2005), no. 1, 189–226. MR2168867

[CF13] Raúl E. Curto and Lawrence A. Fialkow, *Recursively determined representing measures for bivariate truncated moment sequences*, J. Operator Theory **70** (2013), no. 2, 401–436, DOI 10.7900/jot.2011sep06.1943. MR3138363

[dDS18a] Philipp J. di Dio and Konrad Schmüdgen, *The multidimensional truncated moment problem: atoms, determinacy, and core variety*, J. Funct. Anal. **274** (2018), no. 11, 3124–3148, DOI 10.1016/j.jfa.2017.11.013. MR3782989

[dDS18b] Philipp J. di Dio and Konrad Schmüdgen, *The multidimensional truncated moment problem: Carathéodory numbers*, J. Math. Anal. Appl. **461** (2018), no. 2, 1606–1638, DOI 10.1016/j.jmaa.2017.12.021. MR3765506

[Fia17] Lawrence A. Fialkow, *The core variety of a multisequence in the truncated moment problem*, J. Math. Anal. Appl. **456** (2017), no. 2, 946–969, DOI 10.1016/j.jmaa.2017.07.041. MR3688460

[FN10] Lawrence Fialkow and Jiawang Nie, *Positivity of Riesz functionals and solutions of quadratic and quartic moment problems*, J. Funct. Anal. **258** (2010), no. 1, 328–356, DOI 10.1016/j.jfa.2009.09.015. MR2557966

[FP05] Lawrence Fialkow and Srdjan Petrovic, *A moment matrix approach to multivariable cubature*, Integral Equations Operator Theory **52** (2005), no. 1, 85–124, DOI 10.1007/s00020-003-1334-9. MR2138699

[Gru09] Gerd Grubb, *Distributions and operators*, Graduate Texts in Mathematics, vol. 252, Springer, New York, 2009. MR2453959

[IKLS17] M. Infusino, T. Kuna, J. L. Lebowitz, and E. R. Speer, *The truncated moment problem on $\mathbb{N}_0$*, J. Math. Anal. Appl. **452** (2017), no. 1, 443–468, DOI 10.1016/j.jmaa.2017.02.060. MR3628029

[Kem68] J. H. B. Kemperman, *The general moment problem, a geometric approach*, Ann. Math. Statist **39** (1968), 93–122, DOI 10.1214/aoms/1177698508. MR0247645

[KN77] M. G. Kreĭn and A. A. Nudel′man, *The Markov moment problem and extremal problems: Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development*, Translations of Mathematical Monographs, vol. 50, American Mathematical Society, Providence, R.I., 1977. Translated from the Russian by D. Louvish. MR0458081

[KS53] S. Karlin and L. S. Shapley, *Geometry of moment spaces*, Mem. Amer. Math. Soc. **No. 12** (1953), 93. MR0059329

[Las15] Jean Bernard Lasserre, *An introduction to polynomial and semi-algebraic optimization*, Cambridge Texts in Applied Mathematics, Cambridge University Press, Cambridge, 2015. MR3469431

[Lau09] Monique Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging applications of algebraic geometry, IMA Vol. Math. Appl., vol. 149, Springer, New York, 2009, pp. 157–270, DOI 10.1007/978-0-387-09686-5_7. MR2500468

[Mar08] Murray Marshall, *Positive polynomials and sums of squares*, Mathematical Surveys and Monographs, vol. 146, American Mathematical Society, Providence, RI, 2008. MR2383959

[Mat92] J. Matzke, *Mehrdimensionale Momentenprobleme und Positivitätskegel*, Ph.D. thesis, University of Leipzig, 1992.

[Mer73] Robert C. Merton, *Theory of rational option pricing*, Bell J. Econom. and Management Sci. **4** (1973), 141–183. MR0496534

[MMR05] Jean-Michel Marin, Kerrie Mengersen, and Christian P. Robert, *Bayesian modelling and inference on mixtures of distributions*, Bayesian thinking: modeling and computation, Handbook of Statist., vol. 25, Elsevier/North-Holland, Amsterdam, 2005, pp. 459–507, DOI 10.1016/S0169-7161(05)25016-2. MR2490536

[Pea94] K. Pearson, *Contributions to the mathematical theory of evolution*, Phil. Trans. Roy. Soc. London A **185** (1894), 71–110.

[PFJ06] H. Permuter, J. Francos, and I. Jermyn, *A study of Gaussian mixture models of color and texture features for image classification and segmentation*, Pattern Recognit. **39** (2006), 695–706.

[PS08] Mihai Putinar and Konrad Schmüdgen, *Multivariate determinateness*, Indiana Univ. Math. J. **57** (2008), no. 6, 2931–2968, DOI 10.1512/iumj.2008.57.3692. MR2483007

[Rez92] Bruce Reznick, *Sums of even powers of real linear forms*, Mem. Amer. Math. Soc. **96** (1992), no. 463, viii+155, DOI 10.1090/memo/0463. MR1096187

[Ric57] Hans Richter, *Parameterfreie Abschätzung und Realisierung von Erwartungswerten* (German), Bl. Deutsch. Ges. Versicherungsmath. **3** (1957), 147–162. MR0088130

[Roc72] R. Tyrrell Rockafellar, *Convex analysis*, Princeton Mathematical Series, No. 28, Princeton University Press, Princeton, N.J., 1970. MR0274683

[Rog58] W. W. Rogosinski, *Moments of non-negative mass*, Proc. Roy. Soc. London Ser. A **245** (1958), 1–27, DOI 10.1098/rspa.1958.0062. MR0149217

[RS18] Cordian Riener and Markus Schweighofer, *Optimization approaches to quadrature: new characterizations of Gaussian quadrature on the line and quadrature with few nodes on plane algebraic curves, on the plane and in higher dimensions*, J. Complexity **45** (2018), 22–54, DOI 10.1016/j.jco.2017.10.002. MR3748596

[Sag94] Hans Sagan, *Space-filling curves*, Universitext, Springer-Verlag, New York, 1994. MR1299533

[Sar42] Arthur Sard, *The measure of the critical values of differentiable maps*, Bull. Amer. Math. Soc. **48** (1942), 883–890, DOI 10.1090/S0002-9904-1942-07811-6. MR0007523

[Sch91] Konrad Schmüdgen, *The K-moment problem for compact semi-algebraic sets*, Math. Ann. **289** (1991), no. 2, 203–206, DOI 10.1007/BF01446568. MR1092173

[Sch03] Konrad Schmüdgen, *On the moment problem of closed semi-algebraic sets*, J. Reine Angew. Math. **558** (2003), 225–234, DOI 10.1515/crll.2003.040. MR1979186

[Sch15] Konrad Schmüdgen, *The multi-dimensional truncated moment problem: maximal masses*, Methods Funct. Anal. Topology **21** (2015), no. 3, 266–281. MR3521697

[Sch17] Konrad Schmüdgen, *The moment problem*, Graduate Texts in Mathematics, vol. 277, Springer, Cham, 2017. MR3729411

[SdD17] K. Schmüdgen and P. J. di Dio, *Truncated Moment Problem: Set of Atoms and Carathéodory Numbers*, Oberwolfach Reports **14** (2017), 77–79.

[Sim98] Barry Simon, *The classical moment problem as a self-adjoint finite difference operator*, Adv. Math. **137** (1998), no. 1, 82–203, DOI 10.1006/aima.1998.1728. MR1627806

[Sto16] J. Stoyanov, *Moment properties of probability distributions used in stochastic financial models*, Recent Advances in Financial Engineering 2014 Proceedings of the TMU Finance Workshop 2014, World Scientific Publishing Co. Pte. Ltd., 2016, pp. 1–27.

[TSM85] D. M. Titterington, A. F. M. Smith, and U. E. Makov, *Statistical analysis of finite mixture distributions*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons, Ltd., Chichester, 1985. MR838090

Mathematisches Institut, Universität Leipzig, Augustusplatz 10/11, D-04109 Leipzig, Germany

Max Planck Institute for Mathematics in the Sciences, Inselstrasse 22, D-04103 Leipzig, Germany
Email address: didio@uni-leipzig.de