The Theory of Moments and Applications

Contents

I. Classical Theory **4**

1. Introduction **4**
1.1. Administration 4
1.2. What is the Lecture about 4
1.3. Literature for the Lecture 5

2. Integral Representations of Linear Functionals **5**
2.1. Moment Functionals 5
2.2. Choquet's Theory and adapted Spaces 7
2.3. Existence of Integral Representations 8

3. Moment Problems on Intervals $I \subseteq \mathbb{R}$ **10**
3.1. Moment Sequences, Riesz Functionals, and Hankel Matrices 10
3.2. Hamburger Moment Problem ($I = \mathbb{R}$) 12
3.3. Stieltjes Moment Problem ($I = [0, \infty)$) 12
3.4. Hausdorff Moment Problem ($I = [0, 1]$) 13

4. Multidimensional Moment Problems **14**
4.1. Uniqueness on Compact Sets 14
4.2. Hilbert's Theorem 15
4.3. Examples of Non-negative Polynomials which are not Sums of Squares . . 15
4.4. Schmüdgen's Theorem and Positivstellensatz 17

5. Truncated Moment Problems **17**
5.1. Moment Cone 17
5.2. Supporting Hyperplanes 18
5.3. Richter's Theorem 19

6. Carathéodory Numbers **21**
6.1. Definition of $\mathcal{C}_\mathsf{A}(s)$ and basic Properties 22
6.2. The Carathéodory number for $L : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$ on $\mathcal{X} = \mathbb{R}$ 24
6.3. A Lower Bound for differentiable Functions 25
6.4. Lower Bounds on $\mathcal{C}(L)$ for $L : \mathbb{R}[x, y]_{\leq 2d} \to \mathbb{R}$ on $\mathcal{X} \subseteq \mathbb{R}^2$ open 26
6.5. Lower Bounds on $\mathcal{C}(L)$ for $L : \mathbb{R}[x_1, \dots, x_n]_{\leq d} \to \mathbb{R}$ on $\mathcal{X} \subseteq \mathbb{R}^n$ 27
6.6. $\{s_\mathsf{A}(x) \,|\, x \in \mathcal{X}\}$ Countable 30

II. Applications and More 31

7. Finding atomic representing Measures 31
7.1. Flat Extension . . . 31
7.2. The generalized Eigenvalue Problem and Finding Atomic Representing Measures for $L : \mathbb{R}[x_1,\dots,x_n]_{\leq 2d} \to \mathbb{R}$ on $\mathcal{X} = \mathbb{R}^n$ . . . 34

8. Waring Decomposition 36
8.1. The Apolar Scalar Product and Powers of Linear Forms . . . 36
8.2. The Apolar Scalar Product and Differential Operators . . . 38
8.3. Moment Functionals and Waring Decomposition . . . 39
8.4. The Carathéodory Number and the Waring Rank . . . 42

9. Tensor Decomposition 43
9.1. Definition . . . 43
9.2. Decomposition of Symmetric $2d$-Tensors . . . 43

10. Derivatives of Moments and Moment Functionals 43
10.1. Derivatives of Linear Functionals and Measures . . . 44
10.2. Polytope Reconstruction . . . 45

11. Gaussian Distributions and Mixtures 49
11.1. Reconstruction of One Component . . . 49
11.2. Lower Bounds on Components in a Gaussian Mixture from finitely many Moments . . . 52

12. Moments and Partial Differential Equation 52
12.1. The Heat Equation acting on Moment Sequences: Gaussian Mixtures . . . 52
12.2. The Heat Equation acting on Polynomials . . . 61
12.3. Burgers' Equation . . . 67

13. Transformations of Moment Functionals 69
13.1. Souslin Sets, Isomorphisms, and Lebesgue–Rohlin Spaces . . . 70
13.2. Transformations of Moment Functionals . . . 74

A. Appendix 84
A.1. Hahn–Banach dominated Extension Theorem . . . 84
A.2. The Riesz–Markov–Kakutani Representation Theorem . . . 84
A.3. Stone–Weierstraß Theorem . . . 84
A.4. Carathéodory's Theorem . . . 85
A.5. Sard's Theorem . . . 85
A.6. Daniell's Representation Theorem . . . 85

Index 87

Literature **92**

Part I. Classical Theory

1. Introduction

This is the script for the lecture and therefore not without errors and mistakes. Don't take everything as it is but think about it yourself!

1.1. Administration

Lecture times:

- Wed 10:00 - 11:30, D404
- Thu, 13:30 - 15:00, D404

Exercise time:

- 15 Nov. 13:30 - 15:00, D404
- 29 Nov. 13:30 - 15:00, D404
- 13 Dez. 13:30 - 15:00, D404
- ...

Exams:

- 1/2 Semester:
- Full Semester:

1.2. What is the Lecture about

Moments (name from physics):

$$\int_{\mathbb{R}^3} (x^2+y^2)\cdot\rho(x,y,z)\ \mathrm{d}x. \qquad \text{(Moment of inertia)}$$

Moment Problem: Given a linear space $\mathcal{V}$ of real function $f:\mathcal{X}\to\mathbb{R}$, $\mathcal{X}$ a measurable space, and a linear functional $L:\mathcal{V}\to\mathbb{R}$. Does there exist a measure μ on $\mathcal{X}$ such that

$$L(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$$

for all $f\in\mathcal{V}$?

In most cases we will deal with $V = \mathbb{R}[x_1, \dots, x_n]$ the ring of polynomials and $\mathcal{X} \subseteq \mathbb{R}^n$ a closed bounded or even semi-algebraic set.

We are interested in special measures, e.g., when $\mathcal{V}$ is finite dimensional, then we will have Richter's Theorem: We can chose an finitely atomic measure $\sum_{i=1}^k c_i \cdot \delta_{x_i}$ for L with $k \leq \dim \mathcal{V}$, $c_i > 0$, and $x_i \in \mathcal{X}$ pairwise distinct. Of special theoretical and application interest there is the minimal number k.

Other special measures of interest will be Gaussian mixtures $\sum_{i=1}^k c_i \cdot G(\sigma_i, x_i)$, $G(\sigma, x)$ a Gaussian distribution with variance σ centered at x.

We will also have a look at an optimization point of view:

$$\min_{x_0 \in K \subseteq \mathbb{R}^n} p(x_0) = \min_{\substack{\mu: \operatorname{supp} \mu \subseteq K, \\ \mu(K)=1}} \int p(x)\ \mathrm{d}\mu(x) = \min_{\substack{s\ K\text{-moment} \\ \text{sequece},\ s_0=1}} L_s(p).$$

We hopefully will also be able to study special partial differential equations and measure transformations.

The aim of this lecture is to give a quick introduction in approx. the first half of the semester to the "old" theory and go quickly to recent research.

1.3. Literature for the Lecture

Literature to the Moment Problem and some Applications:

- K. Schmüdgen [Sch17]
- M. Marshall [Mar08]
- M. Laurent [Lau09]
- J.-B. Lasserre [Las15]

Historical literature: [KN77], [Akh65], [AK62].
Literature for convex geometry: [Roc72], [Sch14], [Sim11], [Sol15].
More specialized literature, especially research papers, are cited when needed.

2. Integral Representations of Linear Functionals

2.1. Moment Functionals

Definition 2.1.1. Let $\mathcal{X}$ be a locally compact Hausdorff space and $E \subseteq C(\mathcal{X}, \mathbb{R})$ be a linear subset.

(i) We denote by $\mathcal{M}(\mathcal{X})$ the set of all *Radon measures*.[1]

[1] Note, for us measures are always non-negative, unless specifically denoted as *signed* measure.

(ii) Let $C \subseteq E$ be a subset and $L : E \to \mathbb{R}$ be a linear functional. We call L to be *C-positive* if $L(f) \geq 0$ for all $f \in C$. We call L to be *strictly C-positive* if $L(f) > 0$ for all $f \in C \setminus \{0\}$.

(iii) $E_+ := \{f \in E \,|\, f(x) \geq 0 \text{ for all } x \in \mathcal{X}\}$.

(iv) Let $\mu \in \mathcal{M}(\mathcal{X})$ and $E \subseteq L^1(\mathcal{X}, \mu)$. Then L_μ is the E_+-positive linear functional

$$L_\mu : E \to \mathbb{R}, \quad f \mapsto \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x).$$

Definition 2.1.2. Let $\mathcal{X}$ be a locally convex topological Hausdorff space and $E \subseteq C(\mathcal{X}, \mathbb{R})$ be a linear subset. A linear functional $L : E \to \mathbb{R}$ is called a *moment functional* if there exists a $\mu \in \mathcal{M}(\mathcal{X})$ with $L_\mu = L$. Any such measure μ is called a *representing measure* of L. The *set $\mathcal{M}_L$ of representing measures* of L is

$$\mathcal{M}_L := \{\mu \in \mathcal{M}(\mathcal{X}) \,|\, L = L_\mu\}.$$

A moment functional L is called *determinate* if it has a unique representing measure, i.e., $\#\mathcal{M}_L = 1$.

Lemma 2.1.3. *Let $\mathcal{X}$ be a locally convex topological Hausdorff space, $E \subseteq C(\mathcal{X}, \mathbb{R})$ be a linear subset, and $L : E \to \mathbb{R}$ be a moment functional. Then $\mathcal{M}_L$ is convex.*

Proof. Let $\mu, \nu \in \mathcal{M}_L$ and $\lambda \in [0, 1]$. Then

$$\begin{aligned} L_{\lambda\mu+(1-\lambda)\nu}(f) &= \int_{\mathcal{X}} f(x)\ \mathrm{d}(\lambda\mu + (1-\lambda)\nu)(x) \\ &= \lambda \cdot \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x) + (1-\lambda) \int_{\mathcal{X}} f(x)\ \mathrm{d}\nu(x) \\ &= \lambda L(f) + (1-\lambda) L(f) \\ &= L(f) \end{aligned}$$

for all $f \in E$. □

Definition 2.1.4. Let $K \subseteq \mathcal{X}$ a closed subset of a locally compact topological Hausdorff space. A linear functional $L : E \to \mathbb{R}$ is called a *K-moment functional* if there exists a measure μ on $\mathcal{X}$ such that $\operatorname{supp} \mu \subseteq K$ and $L = L_\mu$. The set of all such representing measures is denoted by

$$\mathcal{M}_{L,K} := \{\mu \in \mathcal{M}(\mathcal{X}) \,|\, \operatorname{supp} \mu \subseteq K \text{ and } L = L_\mu\}.$$

L is called K-determinate if $\#\mathcal{M}_{L,K} = 1$.

2.2. Choquet's[2] Theory and adapted Spaces

Definition 2.2.1. Let $\mathcal{X}$ be a locally compact topological Hausdorff space and $f, g \in C(\mathcal{X}, \mathbb{R})$. We say g *dominates* f if for every $\varepsilon > 0$ there exists a compact subset $K_\varepsilon \subseteq \mathcal{X}$ such that

$$|f(x)| \leq \varepsilon \cdot |g(x)|$$

for all $x \in \mathcal{X} \setminus K_\varepsilon$.

Lemma 2.2.2. *Let*

$$\mathcal{U} := \{\eta \in C_c(\mathcal{X}, \mathbb{R}) \,|\, 0 \leq \eta(x) \leq 1 \text{ for all } x \in \mathcal{X}\}.$$

For any $f, g \in C(\mathcal{X}, \mathbb{R})$ the following are equivalent:

(i) g dominates f.

(ii) For every $\varepsilon > 0$ there exists $\eta_\varepsilon \in U$ such that

$$|f(x)| \leq \varepsilon \cdot |g(x)| + |f(x)| \cdot \eta_\varepsilon(x)$$

for all $x \in \mathcal{X}$.

(iii) For every $\varepsilon > 0$ there exists $h_\varepsilon \in C_c(\mathcal{X}, \mathbb{R})$ such that

$$|f(x)| \leq \varepsilon \cdot |g(x)| + h_\varepsilon(x)$$

for all $x \in \mathcal{X}$.

Proof. (i)⇒(ii): Choose $\eta_\varepsilon \in \mathcal{U}$ such that $\eta_\varepsilon(x) = 1$ for all $x \in K_\varepsilon$.

(ii)⇒(iii): Set $h_\varepsilon := |f| \cdot \eta_\varepsilon$.

(iii)⇒(i): Since $h_\varepsilon \in C_c(\mathcal{X}, \mathbb{R})$ we have that $K_\varepsilon := \operatorname{supp} h_\varepsilon$ is compact and we have $|f(x)| \leq \varepsilon \cdot |g(x)|$ for all $x \in \mathcal{X} \setminus K_\varepsilon$. □

Definition 2.2.3. Let $\mathcal{X}$ be locally compact topological Hausdorff space and $E \subseteq C(\mathcal{X}, \mathbb{R})$ be a linear subspace. E is called *adapted* if the following hold:

(i) $E = E_+ - E_+$.

(ii) For each $x \in \mathcal{X}$ there exists an $f \in E_+$ such that $f(x) > 0$.

(iii) For each $f \in E_+$ there exists an $g \in E_+$ such that g dominates f.

Lemma 2.2.4. *Let $\mathcal{X}$ be a locally compact topological Hausdorff space. If E is an adapted subspace of $C(\mathcal{X}, \mathbb{R})$, then for any $f \in C_c(\mathcal{X}, \mathbb{R})_+$ there exists a $g \in E_+$ such that $g(x) \geq f(x)$ for all $x \in \mathcal{X}$.*

Proof. Let $x \in \mathcal{X}$. By Definition 2.2.3(ii) there exists a $g_x \in E_+$ such that $g_x(x) > 0$. By multiplying g_x with some constant we can assume without loss of generality that

$$g_x(x) > f(x). \tag{$*$}$$

By continuity $(*)$ holds on some neighborhood of x. By compactness of $\operatorname{supp} f$ there are finitely many $x_1, \dots, x_k \in \mathcal{X}$ such that $g(x) := g_{x_1}(x) + \dots + g_{x_k}(x) > f(x)$ for all $x \in \operatorname{supp} f$ and $g(x) \geq f(x)$ for all $x \in \mathcal{X}$. □

[2] Gustave Alfred Arthur Choquet (1 March 1915, Solesmes (France) – 14 November 2006, Lyon)

2.3. Existence of Integral Representations

Lemma 2.3.1. *Let $E \subset F$ be a linear subspace of a real vector space F and let $C \subset F$ be a convex cone of F such that $F = E + C$. Then each $(C \cap E)$-positive linear functional $L : E \to \mathbb{R}$ can be extended to a C-positive linear functional $\tilde{L} : F \to \mathbb{R}$.*

Proof. Let $f \in F$. We define

$$q(f) := \inf\{L(g) \mid g \in E,\ g - f \in C\}. \tag{1}$$

Since $F = E + C$, there exists a $g \in E$ such that $-f + g \in C$, so the corresponding set in (1) is non-empty. It is easy to see that q is a sublinear functional and $L(g) = q(g)$ for $g \in E$. Hence, by the Hahn–Banach dominated Extension Theorem A.1.1 there exists an extension $\tilde{L} : F \to \mathbb{R}$ of $L : E \to \mathbb{R}$ such that $\tilde{L}(f) \leq q(f)$ for all $f \in F$.

Let $h \in C$. Setting $g = 0$ and $f = -h$ we have $g - f \in C$, so that $q(-h) \leq L(0) = 0$ by $(*)$. Hence, $\tilde{L}(-h) \leq q(-h) \leq 0$, so $\tilde{L}(h) \geq 0$ and $\tilde{L}$ is C-positive. □

Basic Representation Theorem 2.3.2. *Let $\mathcal{X}$ be a locally compact topological Hausdorff space and $E \subseteq C(\mathcal{X}, \mathbb{R})$ be an adapted subspace. For any linear functional $L : E \to \mathbb{R}$ the following are equivalent:*

(i) The functional L is E_+-positive.

(ii) For each $f \in E_+$ there exists an $h \in E_+$ such that $L(f + \varepsilon h) \geq 0$ for all $\varepsilon > 0$.

(iii) L is a moment functional.

Proof. The implications (iii) $\Rightarrow$ (i) $\Leftrightarrow$ (ii) are clear.

(i) $\Rightarrow$ (iii): Set

$$\tilde{E} := \{f \in C(\mathcal{X}, \mathbb{R}) \mid \text{there exists } g \in E \text{ such that } |f(x)| \leq |g(x)| \text{ for all } x \in \mathcal{X}\}.$$

We show $\tilde{E} = E + (\tilde{E})_+$. We have $E + (\tilde{E})_+ \subseteq \tilde{E}$. Conversely, let $f \in \tilde{E}$. We chose $g \in E_+$ such that $|f| \leq g$. Then we have $f + g \in (\tilde{E})_+$, $-g \in E$, and $f = -g + (g + f) \in E + (\tilde{E})_+$. Hence, $\tilde{E} = E + (\tilde{E})_+$.

By Lemma 2.3.1 we can extend L to an $(\tilde{E})_+$-positive linear functional $\tilde{L} : \tilde{E} \to \mathbb{R}$. We have $C_c(\mathcal{X}, \mathbb{R}) \subseteq \tilde{E}$ by Lemma 2.2.4 and hence by the Riesz–Markov–Kakutani Representation Theorem A.2.1 there exists a representing measure $\mu \in \mathcal{M}(\mathcal{X})$ such that $\tilde{L}(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$ for all $f \in C_c(\mathcal{X}, \mathbb{R})$. By Definition 2.2.3(i) we have $E = E_+ - E_+$, i.e., is remains to show that for all $f \in E_+$ we have $f \in L^1(\mathcal{X}, \mu)$ and $L(f) \equiv \tilde{L}(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$.

Let $f \in E_+$ and set

$$\mathcal{U} := \{\eta \in C_c(\mathcal{X}, \mathbb{R}) \mid 0 \leq \eta(x) \leq 1 \text{ for all } x \in \mathcal{X}\}.$$

For $\eta \in \mathcal{U}$ we have $f \cdot \eta \in C_c(\mathcal{X}, \mathbb{R})$ and hence $\tilde{L}(f \cdot \eta) = \int_{\mathcal{X}} f(x) \cdot \eta(x)\ \mathrm{d}\mu(x)$. From this and the $(\tilde{E})_+$-positivity of $\tilde{L}$ we have

$$\int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x) = \sup_{\eta \in \mathcal{U}} \int_{\mathcal{X}} f(x) \cdot \eta(x)\ \mathrm{d}\mu(x) = \sup_{\eta \in \mathcal{U}} \tilde{L}(f \cdot \eta) \leq \tilde{L}(f) = L(f) < \infty \tag{2}$$

and therefore $f \in L^1(\mathcal{X}, \mu)$.

By (2) it is sufficient to show $L(f) \leq \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$. By Definition 2.2.3(iii) there exists a $g \in E_+$ that dominates f. Then, by Lemma 2.2.2 for any $\varepsilon > 0$ there exists a function $\eta_\varepsilon \in \mathcal{U}$ such that $f \leq \varepsilon \cdot g + f \cdot \eta_\varepsilon$. Since $f \cdot \eta_\varepsilon \leq f$ we obtain

$$L(f) = \tilde{L}(f) \leq \varepsilon \tilde{L}(g) + \tilde{L}(f \cdot \eta_\varepsilon) = \varepsilon L(g) + \int f(x) \cdot \eta_\varepsilon(x)\ \mathrm{d}\mu(x) \leq \varepsilon L(g) + \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x).$$

Since g does not depend on ε we pass to the limit $\varepsilon \searrow 0$ to get $L(f) \leq \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$. Hence, $L(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$ which completes the proof. □

Proposition 2.3.3. *Let $\mathcal{X}$ be a compact Hausdorff space and $E \subseteq C(\mathcal{X}, \mathbb{R})$ be a linear subspace such that there exists an $e \in E$ with $e(x) > 0$ for all $x \in \mathcal{X}$. Then each E_+-positive linear functional $L : E \to \mathbb{R}$ is a moment functional.*

Proof. Set $F = C(\mathcal{X}, \mathbb{R})$ and $C = C(\mathcal{X}, \mathbb{R})_+$. Let $f \in F$. Since $\mathcal{X}$ is compact, f is bounded and e has a positive minimum. Hence, there exists a $\lambda > 0$ such that $f(x) \leq \lambda e(x)$ for all $x \in \mathcal{X}$. Since $\lambda e - f \in C$ and $-\lambda e \in E$ we have

$$-f = -\lambda e + (\lambda e - f) \in E + C,$$

i.e., $F = E + C$. By Lemma 2.3.1 L extends to a C-positive linear functional $\tilde{L} : F \to \mathbb{R}$. By the Riesz–Markov–Kakutani Representation Theorem A.2.1 $\tilde{L}$ and hence also L have a representing measure $\mu \in \mathcal{M}(\mathcal{X})$. □

Definition 2.3.4. Let $n \in \mathbb{N}$ and $K \subseteq \mathbb{R}^n$. Then we define the cone of non-negative polynomials on K by

$$\mathrm{Pos}(K) := \{f \in \mathbb{R}[x_1, \dots, x_n] \mid f(x) \geq 0 \text{ for all } x \in K\}$$

and the cone of sums of squares $\sum \mathbb{R}[x_1, \dots, x_n]^2$ by

$$\sum \mathbb{R}[x_1, \dots, x_n]^2 := \{f_1^2 + \dots + f_d^2 \mid f_1, \dots, f_d \in \mathbb{R}[x_1, \dots, x_n] \text{ for some } d \in \mathbb{N}\}.$$

Haviland's Theorem 2.3.5 ([Hav35, Hav36][3])**.** *Let $n \in \mathbb{N}$, $K \subseteq \mathbb{R}^n$ be a closed subset, and $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$. Then the following are equivalent:*

(i) L is $\mathrm{Pos}(K)$-positive.

(ii) $L(f + \varepsilon \cdot 1) \geq 0$ for all $f \in \mathrm{Pos}(K)$ and $\varepsilon > 0$.

(iii) For any $f \in \mathrm{Pos}(K)$ there is an $h \in \mathrm{Pos}(K)$ such that $L(f + \varepsilon h) \geq 0$ for all $\varepsilon > 0$.

(iv) L is a K-moment functional.

Proof. (iv) ⇒ (i) ⇔ (ii) ⇔ (iii) are clear. It is sufficient to prove (i) ⇒ (iv).

We check that $E = \mathbb{R}[x_1, \dots, x_n] \subseteq C(\mathcal{X} = \mathbb{R}^n, \mathbb{R})$ is an adapted space. Since $4p = (p+1)^2 - (p-1)^2$ condition (i) in Definition 2.2.3 is fulfilled. Condition (ii) of Definition 2.2.3 is fulfilled since $1 \in E_+ = \mathrm{Pos}(K)$ and condition (iii) is fulfilled since for any $p \in E_+ = \mathrm{Pos}(K)$ we have that $g := (x_1^2 + \dots + x_n^2) \cdot p$ dominates p. Now the Basic Representation Theorem 2.3.2 applies and proves the statement. □

[3] Edward Kenneth Haviland (1934 PhD Johns Hopkins University)

3. Moment Problems on Intervals $I \subseteq \mathbb{R}$

3.1. Moment Sequences, Riesz Functionals, and Hankel Matrices

Definition 3.1.1. Let $n \in \mathbb{N}$, $K \subseteq \mathbb{R}^n$ be a closed subset, and let $s = (s_\alpha)_{\alpha \in \mathbb{N}_0^n}$ be a real sequence. We call s a *K-moment sequence* (or just *moment sequence*) if there exists a measure $\mu \in \mathcal{M}(\mathbb{R}^n)$ with $\operatorname{supp}\mu \subseteq K$ and

$$s_\alpha = \int_K x^\alpha \, \mathrm{d}\mu(x)$$

for all $\alpha \in \mathbb{N}_0^n$. s_α is called the *αth moment*.

Definition 3.1.2. Let $n \in \mathbb{N}$ and $s = (s_\alpha)_{\alpha \in \mathbb{N}_0^n}$ be a real sequence. We define the *Riesz functional* $L_s : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ by $L_s(x^\alpha) := s_\alpha$ for all $\alpha \in \mathbb{N}_0^n$ and extend it linearly to all $\mathbb{R}[x_1, \dots, x_n]$.

Example 3.1.3. Let $s = (s_i)_{i \in \mathbb{N}_0}$ be a real sequence and $p(x) = \sum_{i=0}^d c_i x^i \in \mathbb{R}[x]$. Then

$$L_s(p) = \sum_{i=0}^d c_i s_i. \qquad \circ$$

Lemma 3.1.4. *Let $n \in \mathbb{N}$, $K \subseteq \mathbb{R}^n$ be closed, and $s = (s_\alpha)_{\alpha \in \mathbb{N}_0^n}$ be a real sequence. The following are equivalent:*

(i) s is a K-moment sequence.

(ii) L_s is a K-moment functional.

Proof. Follows directly from the definitions. □

Definition 3.1.5. Let $n \in \mathbb{N}$ and $s = (s_\alpha)_{\alpha \in \mathbb{N}_0^n}$ be a real sequence. For each $d \in \mathbb{N}_0$ we define the *Hankel*[4] *matrix* $H_d(s)$ by

$$H_d(s) = (s_{\alpha+\beta})_{\alpha,\beta \in \mathbb{N}_0^n : |\alpha|, |\beta| \leq d} \quad \in \mathbb{R}^{N \times N}$$

with $|\alpha| := \alpha_1 + \dots + \alpha_n$ for all $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$ and $N = \binom{n+d}{n}$.

Example 3.1.6. Let $s = (s_i)_{i \in \mathbb{N}_0}$ be a real sequence and $d \in \mathbb{N}_0$. Then

$$H_d(s) = \begin{pmatrix} s_0 & s_1 & s_2 & \cdots & s_d \\ s_1 & s_2 & s_3 & \cdots & s_{d+1} \\ s_2 & s_3 & s_4 & \cdots & s_{d+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_d & s_{d+1} & s_{d+2} & \cdots & s_{2d} \end{pmatrix}. \qquad \circ$$

[4] Hermann Hankel (14 February 1839, Halle (Saale) – 29 August 1873, Schramberg)

Definition 3.1.7. Let $n \in \mathbb{N}$, $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n}$ be a real sequence, and $\gamma \in \mathbb{N}_0^n$. We define the *shift* X^γ acting on s by

$$X^\gamma s = (s_{\alpha+\gamma})_{\alpha\in\mathbb{N}_0^n}.$$

Example 3.1.8. Let $s = (s_i)_{i\in\mathbb{N}_0}$ be a real sequence and $k, d \in \mathbb{N}_0$. Then

$$H_d(X^k s) = \begin{pmatrix} s_k & s_{k+1} & s_{k+2} & \cdots & s_{k+d} \\ s_{k+1} & s_{k+2} & s_{k+3} & \cdots & s_{k+d+1} \\ s_{k+2} & s_{k+3} & s_{k+4} & \cdots & s_{k+d+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{k+d} & s_{k+d+1} & s_{k+d+2} & \cdots & s_{k+2d} \end{pmatrix}. \qquad \circ$$

Definition 3.1.9. Let $n \in \mathbb{N}$ and $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n}$ be a real sequences. We call s a *positive semidefinite sequence* if $H_d(s) \succeq 0$ for all $d \in \mathbb{N}_0$, i.e.,

$$x^T \cdot H_d(s) \cdot x = \sum_{\alpha,\beta\in\mathbb{N}_0^n:|\alpha|,|\beta|\le d} x_\alpha \cdot s_{\alpha+\beta} \cdot x_\beta \ge 0$$

for all $x = (x_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\le d} \in \mathbb{R}^N$ with $N = \binom{n+d}{n}$. We call s a *positive definite sequence* if $H_d(s) \succ 0$ for all $d \in \mathbb{N}_0$, i.e.,

$$x^T \cdot H_d(s) \cdot x = \sum_{\alpha,\beta\in\mathbb{N}_0^n:|\alpha|,|\beta|\le d} x_\alpha \cdot s_{\alpha+\beta} \cdot x_\beta > 0$$

for all $x = (x_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\le d} \in \mathbb{R}^N \setminus \{0\}$ with $N = \binom{n+d}{n}$.

Lemma 3.1.10. *Let $n \in \mathbb{N}$, $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n}$ be a real sequence, $c = (c_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\le d}$ with $d \in \mathbb{N}_0$, and $\gamma \in \mathbb{N}_0^n$. The following holds:*

(i) For $p_c(x) := \sum_{\alpha\in\mathbb{N}_0^n:|\alpha|\le d} c_\alpha x^\alpha \in \mathbb{R}[x_1,\dots,x_n]$ we have

$$L_s(p_c^2) = c^T \cdot H_d(s) \cdot c.$$

(ii) For all $p \in \mathbb{R}[x_1,\dots,x_n]$ we have

$$L_{X^\gamma s}(p) = L_s(x^\gamma \cdot p(x)).$$

Proof. Follows directly from the definitions. $\square$

3.2. Hamburger Moment Problem ($I = \mathbb{R}$)

Lemma 3.2.1. $\mathrm{Pos}(\mathbb{R}) = \{f^2 + g^2 \mid f, g \in \mathbb{R}[x]\}$.

Proof. $\supseteq$ is clear. So let $p \in \mathrm{Pos}(\mathbb{R})$. By the fundamental theorem of algebra and since all coefficients are real we can write p as

$$p(x) = \prod_{i=1}^{k} (x - a_i)^2 \cdot \prod_{j=1}^{l} \underbrace{(x - b_j)(x - \overline{b_j})}_{=(x-c_j)^2+d_j^2}$$

with $a_i \in \mathbb{R}$ for all $i = 1, \dots, k$ and $b_j \in \mathbb{C} \setminus \mathbb{R}$ as well as $c_j, d_j \in \mathbb{R}$ for all $j = 1, \dots, l$ for some $k, l \in \mathbb{N}_0$ with $2k + 2l = \deg p$. From

$$(a^2 + b^2)(c^2 + d^2) = (ac - bd)^2 + (ad + bc)^2$$

we find $p = f^2 + g^2$ for some $f, g \in \mathbb{R}[x]$ which proves $\subseteq$. □

The following is the solution to the Hamburger[5] moment problem, i.e., $I = \mathbb{R}$.

Hamburger's Theorem 3.2.2 ([Ham20])**.** *Let $s = (s_i)_{i \in \mathbb{N}_0}$ be a real sequence. The following are equivalent:*

(i) s is a $\mathbb{R}$-moment sequence (Hamburger moment sequence).

(ii) $L_s(p) \geq 0$ for all $p \in \mathrm{Pos}(\mathbb{R})$.

(iii) $L_s(p^2) \geq 0$ for all $p \in \mathbb{R}[x]$.

(iv) s is positive semidefinite, i.e., $H_d(s) \succeq 0$ for all $d \in \mathbb{N}_0$.

Proof. From Haviland's Theorem 2.3.5 we have (i) ⇔ (ii). From Lemma 3.2.1 we have (ii) ⇔ (iii). (iii) ⇔ (iv) follows from Lemma 3.1.10. □

3.3. Stieltjes Moment Problem ($I = [0, \infty)$)

Lemma 3.3.1. $\mathrm{Pos}([0,\infty)) = \{f_1(x)^2 + f_2(x)^2 + x \cdot (g_1(x)^2 + g_2(x)^2) \mid f_1, f_2, g_1, g_2 \in \mathbb{R}[x]\}$.

Proof. Set $Q := \{f_1(x)^2 + f_2(x)^2 + x \cdot (g_1(x)^2 + g_2(x)^2) \mid f_1, f_2, g_1, g_2 \in \mathbb{R}[x]\}$. Then $\mathrm{Pos}([0,\infty)) \supseteq Q$ is clear. So let $p \in \mathrm{Pos}([0,\infty))$. By the fundamental theorem of algebra we can write p as

$$p(x) = a \cdot \prod_{i=1}^{k} (x - a_i)^{d_i} \cdot \prod_{j=1}^{l} ((x - b_j)^2 + c_j^2)^{e_j} \tag{$*$}$$

for some $a, a_i, b_j, c_j \in \mathbb{R}$, $d_i, e_j \in \mathbb{N}$ for all $i = 1, \dots, k$, $j = 1, \dots, l$, and $k, l \in \mathbb{N}_0$. By

$$(f_1^2 + x g_1^2)(f_2^2 + x g_2^2) = (f_1^2 f_2^2 + x^2 g_1^2 g_2^2) + x(f_1^2 g_2^2 + g_1 f_2^2) \in Q$$

[5]Hans Ludwig Hamburger (5 August 1889, Berlin – 14 August 1956, Cologne)

we have $Q\cdot Q \subseteq Q$. Hence, in $(*)$ it is sufficient to to show that every factor is in Q. Here, $(x-a_i)^{d_i}, ((x-b_j)^2+c_j^2)^{e_j} \in Q$ for all $j=1,\dots,l$ and even d_i is clear. So let us look at $a \in \mathbb{R}$ and $(x-a_i)^{d_i}$ with d_i odd, assume $d_i = 1$ since $d_i = 2\delta_i + 1$ implies $(x-a_i)^{2\delta_i} \in Q$ again. From $\lim_{x\to\infty} p(x) = \infty$ we gain $a > 0$ and since $a_i \in \mathbb{R}$ are disjoint, at every $x = a_i$ the polynomial $p(x)$ has a sign change, i.e., $a_i \leq 0$ and $x - a_i = -a_i + x \in Q$. $\square$

The following is the solution to the Stieltjes[6] moment problem, i.e., $I = [0,\infty)$.

Stieltjes' Theorem 3.3.2 ([Sti94]). *Let $s = (s_i)_{i\in\mathbb{N}_0}$ be a real sequence. The following are equivalent:*

(i) s is a $[0,\infty)$-moment sequence (Stieltjes moment sequences).

(ii) $L_s(p) \geq 0$ for all $p \in \operatorname{Pos}([0,\infty))$.

(iii) $L_s(p^2) \geq 0$ and $L_{Xs}(p^2) \geq 0$ for all $p \in \mathbb{R}[x]$.

(iv) s and Xs are positive semidefinite, i.e., $H_d(s) \succeq 0$ and $H_d(Xs) \succeq 0$ for all $d \in \mathbb{N}_0$.

Proof. From Haviland's Theorem 2.3.5 we have (i) $\Leftrightarrow$ (ii). From Lemma 3.3.1 we have (ii) $\Leftrightarrow$ (iii). (iii) $\Leftrightarrow$ (iv) follows from Lemma 3.1.10. $\square$

The next example gives the first indeterminate moment sequence.

Example 3.3.3 ([Sti94]). Let $c \in [-1,1]$ and

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot \chi_{(0,\infty)}(x) \cdot x^{-1} \cdot \exp\left(-\frac{1}{2}(\ln x)^2\right)$$

for all $x \in \mathbb{R}$ (or $x \in [0,\infty)$). Then the measure $\mu_c \in \mathcal{M}(\mathbb{R})$ defined by

$$\mathrm{d}\mu_c(x) := [1 + c\cdot\sin(2\pi\ln x)]\cdot f(x)\ \mathrm{d}x$$

has the moments

$$s_k = \int_{\mathbb{R}} x^k\ \mathrm{d}\mu_c(x) = e^{\frac{1}{2}k^2}$$

for all $k \in \mathbb{N}_0$, i.e., independent on $c \in [-1,1]$. $\circ$

3.4. Hausdorff Moment Problem ($I = [0,1]$)

Lemma 3.4.1. $\operatorname{Pos}([0,1]) = \{f(x) + x\cdot g(x) + (1-x)\cdot h(x) \mid f,g,h \in \sum\mathbb{R}[x]^2\}$.

Proof. Set $Q = \{f(x) + x\cdot g(x) + (1-x)\cdot h(x) \mid f,g,h \in \sum\mathbb{R}[x]^2\}$. Then $Q \subseteq \operatorname{Pos}([0,1])$ is clear. So let $p \in \operatorname{Pos}([0,1])$. Then by the fundamental theorem of algebra we can write p as

$$p(x) = a\cdot\prod_{i=1}^{k}(x-a_i)^{d_i}\cdot\prod_{j=1}^{l}((x-b_j)^2+c_j^2)^{e_j} \qquad (*)$$

[6] Thomas Joannes Stieltjes (29 December 1856, Zwolle (Netherlands) – 31 December 1894, Toulouse)

for some $a, a_i, b_j, c_j \in \mathbb{R}$, $d_i, e_j \in \mathbb{N}$ for all $i = 1, \dots, k$, $j = 1, \dots, l$, and $k, l \in \mathbb{N}_0$. From

$$x \cdot (1-x) = x^2 \cdot (1-x) + x \cdot (1-x)^2$$

we see that $Q \cdot Q \subseteq Q$. Let us look at $(x-a_i)^{d_i}$ for d_i odd and we can assume $d_i = 1$ since with $d_i = 2\delta_i + 1$ we have $(x-a_i)^{\delta_i} \in Q$. But since the a_i are disjoint, we have at a_i a sign change, i.e., $a_i \notin (0,1)$. Since $p(x) \geq 0$ on $[0,1]$ we then have that we can change the sign of a and for all $a_i > 1$ the sign from $(x - a_i)$ to $(a_i - x)$ to have $p \in Q$. □

The following is the solution to the Hausdorff[7] moment problem, i.e., $I = [0,1]$.

Hausdorff's Theorem 3.4.2 ([Hau21]). *Let $s = (s_i)_{i\in\mathbb{N}_0}$ be a real sequence. The following are equivalent:*

(i) s is a $[0,1]$-moment sequence (Hausdorff moment sequences).

(ii) $L_s(p) \geq 0$ for all $p \in \operatorname{Pos}([0,1])$.

(iii) $L_s(p^2) \geq 0$, $L_{Xs}(p^2) \geq 0$, and $L_{(1-X)s}(p^2) \geq 0$ for all $p \in \mathbb{R}[x]$.

(iv) s, Xs, and $(1-X)s$ are positive semidefinite, i.e., $H_d(s) \succeq 0$, $H_d(Xs) \succeq 0$, and $H_d((1-X)s) \succeq 0$ for all $d \in \mathbb{N}_0$.

Proof. From Haviland's Theorem 2.3.5 we have (i) ⇔ (ii). From Lemma 3.4.1 we have (ii) ⇔ (iii). (iii) ⇔ (iv) follows from Lemma 3.1.10. □

4. Multidimensional Moment Problems

4.1. Uniqueness on Compact Sets

Theorem 4.1.1. *Let $n \in \mathbb{N}$ and $K \subset \mathbb{R}^n$ be compact. Then every $\operatorname{Pos}(K)$-linear functional $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ has a unique representing measure, i.e., L is determinate.*

Proof. The existence of a representing measure follows from Haviland's Theorem 2.3.5. It remains to show the uniqueness. Since $\mathbb{R}[x_1, \dots, x_n] \subseteq C(K, \mathbb{R})$ is a unitial algebra which separates points, by the Stone–Weierstraß Theorem A.3.1 the polynomials are dense in $C(K, \mathbb{R})$. Now assume L has two representing measures μ_1 and μ_2 such that $\mu_1 \neq \mu_2$. Then by the Riesz–Markov–Kakutani Representation Theorem A.2.1 there exists a $f \in C_c(K, \mathbb{R}) = C(K, \mathbb{R})$ with

$$\int_K f(x)\ \mathrm{d}\mu_1(x) \neq \int_K f(x)\ \mathrm{d}\mu_2(x). \tag{$*$}$$

Since $\mathbb{R}[x_1, \dots, x_n]$ is dense in $C(K, \mathbb{R})$ there exists a family $(p_k)_{k\in\mathbb{N}} \subseteq \mathbb{R}[x_1, \dots, x_n]$ with

$$\|p_k - f\|_\infty \leq \frac{1}{k}$$

[7]Felix Hausdorff (8 November 1868, Breslau – 26 January 1942, Bonn)

for all $k \in \mathbb{N}$. Then

$$\begin{aligned}\left|\int f(x)\,\mathrm{d}\mu_1(x) - \int f(x)\,\mathrm{d}\mu_2(x)\right| &\leq \left|\int f(x) - p_k(x)\,\mathrm{d}\mu_1(x)\right| + \left|\int f(x) - p_k(x)\,\mathrm{d}\mu_2(x)\right| \\ &\leq \|f - p_k\|_\infty \cdot \left(\left|\int 1\,\mathrm{d}\mu_1(x)\right| + \left|\int 1\,\mathrm{d}\mu_2(x)\right|\right) \\ &\leq \frac{2}{k}\cdot L(1) \quad \xrightarrow{k\to\infty} \quad 0\end{aligned}$$

which contradicts $(*)$ and therefore $\mu_1 = \mu_2$. □

4.2. Hilbert's[8] Theorem

Definition 4.2.1. Let $n \in \mathbb{N}$ and $d \in \mathbb{N}_0$. We define

$$\Sigma(n,d) := \Sigma\mathbb{R}[x_1,\dots,x_n]^2 \cap \mathbb{R}[x_1,\dots,x_n]_{\leq d}$$

and

$$\mathrm{Pos}(n,d) := \mathrm{Pos}(\mathbb{R}^n) \cap \mathbb{R}[x_1,\dots,x_n]_{\leq d}.$$

Hilbert's Theorem 4.2.2 ([Hil88])**.** *We have*

$$\mathrm{Pos}(n,d) = \Sigma(n,d) \qquad \Leftrightarrow \qquad (n,d) \in \quad \{1\}\times 2\mathbb{N}_0 \quad \cup \quad \mathbb{N}\times\{2\} \quad \cup \quad \{(2,4)\}.$$

4.3. Examples of Non-negative Polynomials which are not Sums of Squares

Definition 4.3.1. We define the *Motzkin*[9] *polynomial* by

$$f_{\mathrm{Motzkin}}(x,y) := 1 - 3x^2y^2 + x^2y^4 + x^4y^2.$$

Theorem 4.3.2 ([Mot67])**.** $f_{\mathrm{Motzkin}} \in \mathrm{Pos}(2,6) \setminus \Sigma(2,6)$.

Proof. From the inequality

$$\frac{a+b+c}{3} \geq \sqrt[3]{abc}$$

for $a,b,c \geq 0$ we get with $a = 1$, $b = x^2y^4$ and $c = x^4y^2$ that $f_{\mathrm{Motzkin}}(x,y) \in \mathrm{Pos}(2,6)$.

We now show $f_{\mathrm{Motzkin}} \notin \Sigma(2,6)$. Assume we have $f_{\mathrm{Motzkin}} = \sum_i f_i^2$ for some $f_i \in \mathbb{R}[x,y]$. Then $\deg f_i \leq 3$ and all f_i are linear combinations of 1, x, y, x^2, xy, y^2, x^3, x^2y, xy^2, y^3.

Assume x^3 appears in some f_i. Then also x^6 would appear in f_{Motzkin} and therefore x^3 does not appear in any f_i. The same holds for y^3.

[8] David Hilbert (23 January 1862, Königsberg – 14 February 1943, Göttingen)
[9] Theodore Samuel Motzkin (26 March 1908, Berlin – 15 December 1970, Los Angeles)

Assume x^2 appears in some f_i. Since no x^3 appears and could give a negative coefficient from a possible x, then also x^4 would appear in f_{Motzkin}, i.e., x^2 does not appear in any f_i. The same holds for y^2.

Assume x appears in some f_i. Since no x^2 appears and could give a negative coefficient from a possible 1, then also x^2 would appear in f_{Motzkin}, i.e., x does not appear in any f_i. The same holds for y.

In summary, every f_i must be of the form

$$f_i = a_i + b_i xy + c_i x^2 y + d_i xy^2.$$

But then $\sum_i b_i^2 = -3$ contradicts $f_{\text{Motzkin}} \in \Sigma(2,6)$, i.e., $f_{\text{Motzkin}} \notin \Sigma(2,6)$. □

Examples 4.3.3. **(a)** Robinson polynomial [Rob69]:

$$f_{\text{Robinson}}(x,y) = 1-x^2-y^2-x^4+3x^2y^2-y^4+x^6-x^4y^2-x^2y^4+y^6 \quad \in \operatorname{Pos}(2,6)\setminus\Sigma(2,6).$$

(b) Choi–Lam[10] polynomial [CL77]:

$$f_{\text{Choi–Lam}}(x,y,z) = 1 - 4xyz + x^2y^2 + x^2z^2 + y^2z^2 \quad \in \operatorname{Pos}(3,4) \setminus \Sigma(3,4).$$

(c) Schmüdgen[11] polynomial [Sch79]:

$$\begin{aligned} f_{\text{Schmüdgen}}(x,y) = {}& (y^2 - x^2)x(x+2)[x(x-2) + 2(y^2-4)] \\ & + 200[(x^3-4x)^2 + (y^3-4y)^2] \qquad \in \operatorname{Pos}(2,6) \setminus \Sigma(2,6). \end{aligned}$$

(d) Berg–Christensen–Jensen[12] polynomial [BCJ79]:

$$\begin{aligned} f_{\text{Berg–Christensen–Jensen}}(x,y) &= 1 - x^2y^2 + x^4y^2 + x^2y^4 \quad \in \operatorname{Pos}(2,6) \setminus \Sigma(2,6) \\ &= f_{\text{Motzkin}}(x,y) + 2x^2y^2. \end{aligned}$$

(e) Harris[13] polynomial [Har99, $R_{2,0}$ in Lem. 5.1 and 6.8]:

$$\begin{aligned} f_{\text{Harris}}(x,y) = {}& 16x^{10} - 36x^8y^2 + 20x^6y^4 + 20x^4y^6 - 36x^2y^8 + 16y^{10} \\ & - 36x^8 + 57x^6y^2 - 38x^4y^4 + 57x^2y^6 - 36y^8 \\ & + 20x^6 - 38x^4y^2 - 38x^2y^4 + 20y^6 \\ & + 20x^4 + 57x^2y^2 + 20y^4 \\ & - 36x^2 - 36y^2 \\ & + 16 \qquad \in \operatorname{Pos}(2,10) \setminus \Sigma(2,10). \end{aligned}$$

[10] Choi ???
Lam ???

[11] Konrad Schmüdgen (born 11 November 1947, Gräfendorf (Saxony))

[12] Berg ???
Christensen ???
Jensen ???

[13] Harris ???

4.4. Schmüdgen's Theorem and Positivstellensatz

Definition 4.4.1. Let $S = \{g_1, \dots, g_s\} \subset \mathbb{R}[x_1, \dots, x_n]$ be finite, $n, s \in \mathbb{N}$. We denote by K_S the *basic closed semi-algebraic set* by

$$K_S := \{x \in \mathbb{N} \,|\, g_i(x) \geq 0 \text{ for all } i = 1, \dots, s\}.$$

We denote by T_S the preordering[14]

$$T_S := \left\{ \sum_{e \in \{0,1\}^s} \sigma_e \cdot g^e \,\middle|\, \sigma_e \in \Sigma\mathbb{R}[x_1, \dots, x_n]^2 \right\}.$$

Schmüdgen's Theorem 4.4.2 ([Sch91])**.** *Let $n \in \mathbb{N}$, $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ be a linear functional, and $S = \{g_1, \dots, g_s\} \subseteq \mathbb{R}[x_1, \dots, x_n]$ for some $s \in \mathbb{N}$ such that $K_S \subset \mathbb{R}^n$ is compact. The following are equivalent:*

(i) L is a K_S-moment functional.

(ii) $L(g_1^{e_1} \cdots g_s^{e_s} \cdot h^2) \geq 0$ for all $h \in \mathbb{R}[x_1, \dots, x_n]$ and $e_1, \dots, e_s \in \{0, 1\}$.

Schmüdgen's Positivstellensatz 4.4.3. *Let $n \in \mathbb{N}$ and $S \subset \mathbb{R}[x_1, \dots, x_n]$ be finite such that K_S is compact. Then for any $f \in \mathbb{R}[x_1, \dots, x_n]$ with $f(x) > 0$ for all $x \in K_S$ we have $f \in T_S$.*

5. Truncated Moment Problems

5.1. Moment Cone

Definition 5.1.1. Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions $f : \mathcal{X} \to \mathbb{R}$ on a measurable space $\mathcal{X}$. Let $\mathsf{A} = \{a_1, \dots, a_d\}$ be a basis of $\mathcal{V}$, $d \in \mathbb{N}$. We define the set $\mathcal{M}_\mathsf{A}$ of A-integrable measures on $\mathcal{X}$ by

$$\mathcal{M}_\mathsf{A} = \mathcal{M}_\mathsf{A}(\mathcal{X}) := \{\mu \in \mathcal{M}(\mathcal{X}) \,|\, a_1, \dots, a_d \text{ are } \mu\text{-integrable}\}.$$

Lemma 5.1.2. *Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions on a measurable space $\mathcal{X}$ with basis $\mathsf{A} = \{a_1, \dots, a_d\}$, $d \in \mathbb{N}$, such that $a_i(\mathcal{X}) \subseteq \mathbb{R}$. Then $\delta_x \in \mathcal{M}_\mathsf{A}$ for all $x \in \mathcal{X}$.*

Proof. We have

$$\int_\mathcal{X} a_i(y) \, \mathrm{d}\delta_x(y) = a_i(x) \in \mathbb{R}$$

for all $i = 1, \dots, d$. □

[14] $T + T \subseteq T$, $TT \subseteq T$, and $a^2 \in T$ for all $a \in A$

Definition 5.1.3. Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions on a measurable space $\mathcal{X}$ and $\mathsf{A} = \{a_1, \dots, a_d\}$ be a basis of $\mathcal{V}$. Let $a_i(\mathcal{X}) \subseteq \mathbb{R}$ for all $i = 1, \dots, d$. We define the *moment curve* s_A by

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

We define the *moment cone* $\mathcal{S}_\mathsf{A}$ by

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

Lemma 5.1.4. *Let $\mathcal{X}$ be a measurable space, $\mathcal{V}$ be a finite dimensional real vector space of measurable function on $\mathcal{X}$ and $\mathsf{A} = \{a_1, \dots, a_d\}$, $d \in \mathbb{N}$, be a basis of $\mathcal{V}$. Then $\mathcal{S}_\mathsf{A}$ is a convex cone.*

Proof. Clear, since the integral is linear in the measure. □

5.2. Supporting Hyperplanes

Definition 5.2.1. Let $d \in \mathbb{N}$ and $K \subseteq \mathbb{R}^d$ be a convex cone. For $v \in \mathbb{R}^d$ we say that

$$H_v := \{x \in \mathbb{R}^d \,|\, \langle x, v\rangle = 0\}$$

is a *hyperplane* with normal vector v and

$$H_v^+ := \{x \in \mathbb{R}^d \,|\, \langle x, v\rangle \geq 0\}$$

is the corresponding *halfspace*.

We say H_v^+ is a *containing halfspace* iff $\mathcal{S}_\mathsf{A} \subseteq H_v^+$. We call H_v a *supporting hyperplane* iff $\mathcal{S}_\mathsf{A} \subseteq H_v^+$ and $H_v \cap \overline{\mathcal{S}_\mathsf{A}} \neq \emptyset$. Additionally, we say H_v *supports* $\mathcal{S}_\mathsf{A}$ at $s \in \mathcal{S}_\mathsf{A}$ iff H_v is a supporting hyperplane and $s \in H_v$. For $s \in \mathcal{S}_\mathsf{A}$ we define the *normal cone* $\mathrm{Nor}_\mathsf{A}(s)$ by

$$\mathrm{Nor}_\mathsf{A}(s) := \{v \in \mathbb{R}^d \,|\, H_v \text{ is a supporting hyperplane of } \mathcal{S}_\mathsf{A} \text{ at } s\}.$$

Definition 5.2.2. Let $d \in \mathbb{N}$ and $K \subseteq \mathbb{R}^d$ be a convex cone. Then we call K^* defined as

$$K^* := \{v \in \mathbb{R}^d \,|\, K \subseteq H_v^+\}$$

the *dual cone* of K.

Definition 5.2.3. Let K be a convex set. We say a *face* is a convex subset $F \subseteq K$ such that $\lambda x + (1-\lambda)y \in F$ for some $x, y \in K$ and $\lambda \in (0,1)$ implies $x, y \in F$. An *exposed face* F is a face of K such that there exists a hyperplane H with $F = K \cap H$.

Definition 5.2.4. Let $f : \mathcal{X} \to \mathbb{R}$ be a real function. We define the *zero set* $\mathcal{Z}(f)$ by

$$\mathcal{Z}(f) := \{x \in \mathcal{X} \mid f(x) = 0\}.$$

Lemma 5.2.5. *Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions on a measurable space $(\mathcal{X}, \mathfrak{A})$ and A be a basis of $\mathcal{V}$. Let H_v be a supporting hyperplane of $\mathcal{S}_\mathsf{A}$. Then $H_v \cap \mathcal{S}_\mathsf{A}$ is a moment cone on $(\mathcal{Z}, \mathfrak{A}|_\mathcal{Z})$ with $\mathcal{Z} := \mathcal{Z}(\langle v, s_\mathsf{A}(\,\cdot\,)\rangle)$ and $\dim \mathcal{S}_\mathsf{A} \cap H_v < \dim \mathcal{S}_\mathsf{A}$.*

Proof. It is clear that $\dim \mathcal{S}_\mathsf{A} \cap H_v < \dim \mathcal{S}_\mathsf{A}$. By a change in the basis A we can assume that $v = (0, \dots, 0, 1)$ and hence $a_d \geq 0$. Therefore, $s = (s_1, \dots, s_d) \in \mathcal{S}_\mathsf{A} \cap H_v$ implies $s_d = 0$ and

$$0 = s_d = \int a_d(x)\ \mathrm{d}\mu(x)$$

for all $\mu \in \mathcal{M}_\mathsf{A}(s)$ and all $s \in \mathcal{S}_\mathsf{A} \cap H_v$. Then $\mathcal{Z} = \mathcal{Z}(a_d)$ and $\mathfrak{A}|_\mathcal{Z} = \{M \cap \mathcal{Z} \mid M \in \mathfrak{A}\}$. For any $x \in \mathcal{Z}$ we have $a_d(x) \neq 0$, so $s_d \neq 0$, and hence $s_\mathsf{A}(x) \notin \mathcal{S}_\mathsf{A} \cap H_v$. Hence, $\mathcal{S}_\mathsf{A} \cap H_v$ is the moment cone on $(\mathcal{Z}, \mathfrak{A}|_\mathcal{Z})$. □

Proposition 5.2.6. *Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions on a measurable space $\mathcal{X}$. Then there exist point $x_1, \dots, x_d \in \mathcal{X}$ with $d = \dim \mathcal{V}$ such that every vector $s \in \mathbb{R}^d$ has a signed k-atomic representing measure with $k \leq d$ and all atoms are from the set $\{x_1, \dots, x_d\}$.*

Proof. Let A be a basis of $\mathcal{V}$ and since they are linearly independent there are points $x_1, \dots, x_d \in \mathcal{X}$ such that the matrix $(s_\mathsf{A}(x_1), \dots, s_\mathsf{A}(x_d)) \in \mathbb{R}^{d \times d}$ has full rank. Therefore, for any $s \in \mathbb{R}^d$ we have

$$s = c_1 s_\mathsf{A}(x_1) + \dots + c_d s_\mathsf{A}(x_d) = \int s_\mathsf{A}(x)\ \mathrm{d}\left(\sum_{i=1}^d c_i \delta_{x_i}\right)(x)$$

with $(c_1, \dots, c_d) = (s_\mathsf{A}(x_1), \dots, s_\mathsf{A}(x_d))^{-1} s$. □

5.3. Richter's[15] Theorem

Richter's Theorem 5.3.1 ([Ric57])**.** *Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions on a measurable space $\mathcal{X}$. Then any moment functional $L : \mathcal{V} \to \mathbb{R}$ has a k-atomic representing measure with $k \leq \dim \mathcal{V}$.*

Proof. We prove the theorem by induction on the dimension $\dim \mathcal{V}$. Let $\mathsf{A} = \{a_1, \dots, a_d\}$ with $d = \dim \mathcal{V}$ be a basis of $\mathcal{V}$ and $\mathcal{S}_\mathsf{A}$ the moment cone.

Let $d = 1$, i.e., $\mathcal{V} = a \cdot \mathbb{R}$ for an $a \in \mathcal{V} \setminus \{0\}$. Since $L : \mathcal{V} \to \mathbb{R}$ is a moment functional, there exists a representing measure μ:

$$L(a) = \int a(x)\ \mathrm{d}\mu(x).$$

[15] Hans Werner Richter (2 Mai 1912, Schönefeld (Leipzig) – 3 December 1978, Munich)

If $L(a) = 0$, then $L = 0$ and $\nu = 0$ is a 0-atomic representing measure. So assume $L(a) \neq 0$. Since μ is non-negative, there exists a $x \in \mathcal{X}$ such that $\operatorname{sign} a(x) = \operatorname{sign} L(a)$. Hence,

$$L(a) = \frac{L(a)}{a(x)} \cdot a(x) = \int a(y)\ \mathrm{d}\left(\frac{L(a)}{a(x)}\delta_x\right)(y)$$

and therefore $\frac{L(a)}{a(x)} \cdot \delta_x$ is a 1-atomic representing measure of L.

Now let $d > 1$ and $s := (L(a_1), \dots, L(a_d)) \in \mathcal{S}_\mathsf{A}$ be the moment sequence of L. We have $\dim \mathcal{S}_\mathsf{A} = \dim \mathcal{V} = d$. Let $\mathcal{S} := \operatorname{cone} s_\mathsf{A}(\mathcal{X})$, i.e., by Carathéodory's theorem every $s \in \mathcal{S}$ is a moment sequence represented by a k-atomic measure with $k \leq d$. Additionally, we have that $\operatorname{int} \mathcal{S}$ is non-empty, i.e., $\mathcal{S}$ is full dimensional. Assume $\operatorname{int} \mathcal{S} \neq \operatorname{int} \mathcal{S}_\mathsf{A}$, then let $s \in \operatorname{int}(\mathcal{S}_\mathsf{A} \setminus \operatorname{int} \mathcal{S})$ since $\mathcal{S} \subseteq \mathcal{S}_\mathsf{A}$ and the difference set has non-empty interior. Let μ be a representing measure of s. Then there exists a separating linear functional l such that $l(s) < 0$ and $l(t) > 0$ for all $t \in \mathcal{S}$. Since $s_\mathsf{A}(x) \in \mathcal{S} \subset \mathcal{S}_\mathsf{A}$ we have $a(x) := l(s_\mathsf{A}(x)) > 0$ for all $x \in \mathcal{X}$ and hence

$$a > 0 \text{ on } \mathcal{X} \quad \text{but} \quad \int a(x)\ \mathrm{d}\mu(x) = l(s) < 0$$

which contradicts the non-negativity of μ. Hence, we have $\operatorname{int} \mathcal{S} = \operatorname{int} \mathcal{S}_\mathsf{A}$ and every $s \in \operatorname{int} \mathcal{S}_\mathsf{A}$ has a k-atomic representing measure with $k \leq d$.

Now assume $s \in \mathcal{S}_\mathsf{A} \cap \partial \mathcal{S}_\mathsf{A}$. Then since $\mathcal{S}_\mathsf{A}$ is a convex cone there exists a supporting hyperplane H_v at s. But then $\mathcal{S}_\mathsf{A} \cap H_v$ is by Lemma 5.2.5 a moment cone with dimension less than $\dim \mathcal{V}$ and here the theorem holds by induction. $\square$

The replacement of integration by point evaluations was already used and investigated by Gauß[16] [Gau15]. k-atomic representing measures are therefore also called Gaussian cubature formulas.

The history of Richter's Theorem 5.3.1 is confusing and intricate and often the corresponding references in the literature are misleading. For this reason we take this opportunity to discuss this history in detail. First we collect several versions of Richter's Theorem 5.3.1 occurring in the literature in chronological order.

A) A. Wald 1939[17] [Wal39, Proposition 13]: Let $\mathcal{X} = \mathbb{R}$ and $a_i(x) := |x - x_0|^{d_i}$ with $d_i \in \mathbb{N}_0$ and $0 \leq d_1 < d_2 < \cdots < d_m < \infty$ for an $x_0 \in \mathcal{X}$.

B) P. C. Rosenbloom 1952 [Ros52, Corollary 38e]: Let $(\mathcal{X}, \mathfrak{A})$ be a measurable space and a_i bounded measurable functions.

C) H. Richter 1957[18] [Ric57, Satz 4]: Let $(\mathcal{X}, \mathfrak{A})$ be a measurable space and let a_i be measurable functions.

D) M. V. Tchakaloff 1957[19] [Tch57]: Let $\mathcal{X} \subset \mathbb{R}^n$ be compact and $a_\alpha(x) = x^\alpha$, $|\alpha| \leq d$.

[16] Carl Friedrich Gauß (30 April 1777, Braunschweig – 23 Februar 1855, Göttingen)
[17] Received: February 25, 1939. Published: September, 1939.
[18] Received: December 27, 1956. Published: April, 1957.
[19] Published: July-September, 1957.

E) W. W. Rogosinski 1958[20] [Rog58, Theorem 1]: Let $(\mathcal{X}, \mathfrak{A})$ be a measurable space and let a_i be measurable functions.

From this list we see that Tchakaloff's result D) from 1957 is a special case of Rosenbloom's result B) from 1952 and that the general case was proved by Richter and Rogosinski almost about at the same time, see the exact dates in the footnotes. If one reads Richter's paper, one might think at first glance that he treats only the one-dimensional case, but a closer look reveals that his Proposition (Satz) 4 covers actually the general case of measurable functions. Rogosinski treats the one-dimensional case, but he also states that his proof works for general measurable spaces. The above proof of Richter's Theorem 5.3.1, and likewise the one in [Sch17, Theorem 1.24], are nothing but modern formulations of the proofs of Richter and Rogosinski without additional arguments. Note that Rogosinki's paper [Rog58] was submitted about a half year after the appearance of Richter's [Ric57].

It might be of interest to note that the general results of Richter and Rogosinski can be easily derived from Rosenbloom's Theorem by the following simple trick. Let $\mathsf{A} = \{a_1, \dots, a_m\}$ be (finite) measurable functions on $(\mathcal{X}, \mathfrak{A})$ and set $\mathsf{B} = \{b_1, \dots, b_m\}$, where $b_i := \frac{a_i}{f}$ with $f := 1 + \sum_{i=1}^m a_i^2$. Then

$$\begin{aligned} s \in \mathcal{S}_\mathsf{B} \;\Leftrightarrow\; \exists \nu \in \mathfrak{M}_\mathsf{B} : s = \int s_\mathsf{B}(x)\, \mathrm{d}\nu(x) \;\Leftrightarrow\; s = \int \frac{s_\mathsf{A}(x)}{f(x)}\, \mathrm{d}\nu(x) = \int s_\mathsf{A}(x)\, \mathrm{d}\mu(x) \\ \Leftrightarrow\; \exists \mu \in \mathfrak{M}_\mathsf{A} : s = \int s_\mathsf{A}(x)\, \mathrm{d}\mu(x) \;\Leftrightarrow\; s \in \mathcal{S}_\mathsf{A} \quad \text{with} \quad \mathrm{d}\mu = f^{-1}\, \mathrm{d}\nu. \end{aligned}$$

Since all functions b_i are bounded, Rosenbloom's Theorem applies to B, so each sequence $s \in \mathcal{S}_\mathsf{B} = \mathcal{S}_\mathsf{A}$ has a k-atomic representing measure $\nu \in \mathfrak{M}_\mathsf{B}(s)$ with $k \leq m$ and scaling by f^{-1} yields a k-atomic representing measure $\mu \in \mathfrak{M}_\mathsf{A}(s)$:

$$s = \sum_{i=1}^k c_i \cdot s_\mathsf{B}(x_i) = \sum_{i=1}^k \frac{c_i}{f(x_i)} \cdot s_\mathsf{A}(x_i).$$

Richter's Theorem 5.3.1 was overlooked in the modern literature on truncated polynomial moment problems. It was reproved in several papers in weaker forms and finally in the polynomial case in [BT06]. But Richter's Theorem 5.3.1 for general measurable functions was known and cited by J. H. B. Kemperman in [Kem68, Theorem 1] and attributed therein to Richter and Rogosinski. In the moment problem community succeeding Kemperman the general form of Richter's Theorem 5.3.1 was often used, see e.g. [Kem71, eq. (2.3)], [Kem87, page 29], [FP01, Theorem 1, p. 198], [Ana06, Theorem 1], and [Las15, Theorem 2.50].

6. Carathéodory[21]Numbers

[20]Received: August 22, 1957. Published: May 6, 1958.

[21]Constantin Carathéodory (13 September 1873, Berlin – 2 February 1950, Munich)

In this section we treat results on the Carathéodory number which appeared in [Ric57, dDS18, RS18, dDK21, dDS22].

6.1. Definition of $\mathcal{C}_\mathsf{A}(s)$ and basic Properties

Definition 6.1.1. Let $L : \mathcal{V} \to \mathbb{R}$ be a moment functional from a finite dimensional space of real measurable functions $f : \mathcal{X} \to \mathbb{R}$ resp. $s \in \mathcal{S}_\mathsf{A}$, A basis of $\mathcal{V}$, its moment sequence. Let $\mathcal{X}$ be a measurable space. Then

$$\mathcal{C}_\mathsf{A}(s) = \mathcal{C}(L) := \min\{k \,|\, \mu \in \mathcal{M}_\mathsf{A}(s) \text{ is } k\text{-atomic}\}$$

denotes the *Carathéodory number* of s resp. L $(= L_s)$. Let $\mathcal{S}_\mathsf{A}$ be the moment cone, then

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

is the *Carathéodory number* of $\mathcal{S}_\mathsf{A}$.

Definition 6.1.2. Let $\mathcal{V}$ be a finite dimensional real vector space of measurable functions $f : \mathcal{X} \to \mathbb{R}$ and A be a basis of $\mathcal{V}$. For any $k \in \mathbb{N}$ we define the *moment map* $S_{\mathsf{A},k}$ by

$$S_{\mathsf{A},k} : [0,\infty)^k \times \mathcal{X}^k \to \mathbb{R}^{\dim \mathcal{V}}, \quad (c_1,\dots,c_k,x_1,\dots,x_k) \mapsto \sum_{i=1}^k c_i \cdot s_\mathsf{A}(x_i).$$

Additionally, we define $\mathcal{S}_{\mathsf{A},k} := \operatorname{range} S_{\mathsf{A},k}$.

Lemma 6.1.3. *Let $\mathcal{V}$ be a finite dimensional real space of measurable functions $f : \mathcal{X} \to \mathbb{R}$ on a measurable space $\mathcal{X}$ with basis A. The following holds.*

(i) $\mathcal{S}_\mathsf{A} = \operatorname{conv}\operatorname{cone} s_\mathsf{A}(\mathcal{X})$.

(ii) $\mathcal{C}_\mathsf{A} = \min\{k \,|\, \mathcal{S}_{\mathsf{A},k} \textit{ is convex}\}$.

(iii) $\mathcal{S}_{\mathsf{A},0} = \{0\} \subsetneq \mathcal{S}_{\mathsf{A},1} \subsetneq \mathcal{S}_{\mathsf{A},2} \subsetneq \dots \mathcal{S}_{\mathsf{A},\mathcal{C}_\mathsf{A}-1} \subsetneq \mathcal{S}_{\mathsf{A},\mathcal{C}_\mathsf{A}} = \mathcal{S}_{\mathsf{A},\mathcal{C}_\mathsf{A}+1} = \dots$

Proof. (i): Follows immediately from Richter's Theorem 5.3.1.

(ii): Since $s_\mathsf{A}(\mathcal{X}) \subseteq \mathcal{S}_{\mathsf{A},k}$ for all $k \in \mathbb{N}$ we have that

$$\mathcal{S}_\mathsf{A} = \operatorname{conv}\operatorname{cone} s_\mathsf{A}(\mathcal{X}) \subseteq \operatorname{conv}\operatorname{cone} \mathcal{S}_{\mathsf{A},k} = \mathcal{S}_{\mathsf{A},k}$$

for all $k \in \mathbb{N}$ such that $\mathcal{S}_{\mathsf{A},k}$ is convex, i.e., $\mathcal{C}_\mathsf{A} \leq \min\{k \,|\, \mathcal{S}_{\mathsf{A},k} \text{ is convex}\}$. But since $\mathcal{S}_\mathsf{A} = \mathcal{S}_{\mathsf{A},\mathcal{C}_\mathsf{A}}$ is convex, we also have $\mathcal{C}_\mathsf{A} \in \{k \,|\, \mathcal{S}_{\mathsf{A},k} \text{ is convex}\}$.

(iii): Follows immediately from the minimality of $\mathcal{C}_\mathsf{A}$ by its definition. Because if $\mathcal{S}_{\mathsf{A},k} = \mathcal{S}_{\mathsf{A},k+1}$ for some $k \in \mathbb{N}$ we have $\mathcal{S}_{\mathsf{A},k} = \mathcal{S}_{\mathsf{A},l}$ for all $l \geq k$, especially for $l = \dim \mathcal{V} < \infty$ and hence $\mathcal{S}_{\mathsf{A},k} = \mathcal{S}_\mathsf{A}$ and $k \geq \mathcal{C}_\mathsf{A}$. □

Lemma 6.1.4. *Let $\mathcal{V}$ be a finite dimensional real space of measurable functions $f : \mathcal{X} \to \mathbb{R}$ on a measurable space $\mathcal{X}$ with basis A and $a \in \mathcal{V}_+$ such that $\mathcal{Z}(a)$ is finite. Then for*

$$s := \sum_{x\in\mathcal{Z}} s_{\mathsf{A}}(x) \in \mathcal{S}_{\mathsf{A}}$$

we have $L_s(a) = 0$ and

$$\mathcal{C}_{\mathsf{A}}(s) = \dim \operatorname{lin} \{s_{\mathsf{A}}(x) \,|\, x \in \mathcal{Z}(a)\} = \operatorname{rank} (s_{\mathsf{A}}(x))_{x\in\mathcal{Z}(a)} \tag{3}$$

and therefore

$$\mathcal{C}_{\mathsf{A}} \geq \operatorname{rank} (s_{\mathsf{A}}(x))_{x\in\mathcal{Z}(a)}. \tag{4}$$

Proof. The last equality in (3) is clear by linear algebra. By definition of s we have that s is in the relative interior of the face $F \subseteq \mathcal{S}_{\mathsf{A}}$ which is spanned by $s_{\mathsf{A}}(x)$, $x \in \mathcal{Z}(a)$. By Carathéodory's Theorem A.4.1 we have that s is the convex conic linear combination of $\operatorname{rank} (s_{\mathsf{A}}(x))_{x\in\mathcal{Z}}$ extreme points $s_{\mathsf{A}}(x)$, $x \in \mathcal{Z}(a)$. (4) then follows from the definition of $\mathcal{C}_{\mathsf{A}}$. □

Remark 6.1.5. The previous result no longer holds if $\mathcal{Z}(a)$ is infinite. Take e.g. $a = 0$ for an infinite $\mathcal{X}$. ○

$(d-1)$**-Theorem 6.1.6.** *Let $\mathcal{V} \subseteq C(\mathcal{X}, \mathbb{R})$ be a d-dimensional vector space such that there is a $e \in \mathcal{V}_+$ with $e > 0$, $\mathcal{X}$ be a measurable and topological space which consists of at most $d-1$ path-connected components, A be a basis of $\mathcal{V}$, and let $L : \mathcal{V} \to \mathbb{R}$ be a moment functional. Then*

$$\mathcal{C}(L) \leq d - 1$$

and

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

Proof. Since L is a moment functional we have that it has by Richter's Theorem 5.3.1 a k-atomic representing measure $\mu = \sum_{i=1}^k c_i \cdot \delta_{x_i}$ with $k \leq \dim \mathcal{V} = d$. Assume $k = d$ and let s be the moment sequence of L for a basis A of $\mathcal{V}$. W.l.o.g. $\mathsf{A} \cap \mathcal{V}_+ \neq \emptyset$. Then

$$s \in \operatorname{conv} \operatorname{cone} \{s_{\mathsf{A}}(x_1), \dots, s_{\mathsf{A}}(x_d)\}.$$

Since $\mathcal{X}$ has at most $d-1$ path-connected components at least two points x_i belong to the same path connected component of $\mathcal{X}$. W.l.o.g. these two points are x_1 and x_2. Let $\gamma : [0,1] \to \mathcal{X}$ be a continuous path between $x_1 = \gamma(0)$ and $x_2 = \gamma(1)$. Then

$$s \in \operatorname{conv} \operatorname{cone} \{s_{\mathsf{A}}(x_1), s_{\mathsf{A}}(\gamma(t)), s_{\mathsf{A}}(x_3), \dots, s_{\mathsf{A}}(x_d)\}$$

for $t = 1$ and by letting $t \searrow 0$ the convex cone shrinks and degenerates to a $(d-1)$-dimensional cone by continuity of s_{A}. Hence, for some $t \in [0, 1)$ we have that s lies on the boundary of the cone, i.e., it needs only $d-1$ atoms. □

6.2. The Carathéodory number for $L : \mathbb{R}[x]_{\leq d} \to \mathbb{R}$ on $\mathcal{X} = \mathbb{R}$

We denote

$$\lfloor x \rfloor := \max\{k \in \mathbb{Z} \,|\, k \leq x\} \qquad \text{and} \qquad \lceil x \rceil := \min\{k \in \mathbb{Z} \,|\, k \geq x\}.$$

Theorem 6.2.1 ([Ric57]). *Let $\mathcal{X} = \mathbb{R}$, $d \in \mathbb{N}$, $\mathcal{V} = \mathbb{R}[x]_{\leq d}$ with $\mathsf{A} = \{1, \dots, x^d\}$. Then*

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

Proof. (a) Let d be even: Let $s \in \operatorname{int} \mathcal{S}_\mathsf{A}$ be a moment sequence and by Richter's Theorem 5.3.1 let $\mu = \sum_{i=1}^k c_i \cdot \delta_{x_i}$ be a representing measure of s, i.e., $s = \sum_{i=1}^k c_i \cdot s_\mathsf{A}(x_i)$. Let $\mathsf{B} := \{y^d, y^{d-1}x, \dots, x^d\}$ be the homogenization of A, i.e., $\mathcal{V}^h = \mathbb{R}[x,y]_d$ on the projective space $\mathcal{Y} = \mathbb{P}^1$. We set $t := \sum_{i=1}^k c_i \cdot s_\mathsf{B}(x_i, 1) \in \operatorname{int} \mathcal{S}_\mathsf{B}$. Since $\mathcal{Y}$ is compact and $\mathcal{V}^h$ are continuous with $1 \in \mathcal{V}$ we have that $\mathcal{S}_\mathsf{B}$ is pointed and closed. Hence, we have for any $p \in \mathbb{P}^1$ that

$$\lambda_p : \operatorname{int} \mathcal{S}_\mathcal{B} \to (0,\infty), \quad w \mapsto \lambda(w) := \max\{l \geq 0 \,|\, t - l \cdot s_\mathsf{B}(p) \in \mathcal{S}_\mathsf{B}\}$$

is well-defined and we have

$$t' := t - \lambda_{(1,0)}(t) \cdot s_\mathsf{B}(1,0) \in \partial \mathcal{S}_\mathsf{B}$$

and t' has by construction no atom at $(1,0)$. Let U be an open neighborhood of t, then

$$U' := \{u' := u - \lambda_{(1,0)}(u) \cdot s_\mathsf{B}(1,0)\} \subseteq \partial \mathcal{S}_\mathsf{B}$$

is part of the boundary of $\mathcal{S}_\mathsf{B}$ and no $u' \in U'$ has an atomic representing measure with an atom at $(1,0)$. By continuity of B there exist a $\varepsilon > 0$ such that

$$t'' := t - \lambda_{(1,\varepsilon)} \cdot s_\mathsf{B}((1,\varepsilon)) \in U'.$$

Hence, all atomic representing measures ν of t'' have not $(0,1)$ as an atom. Since $t'' \in U' \subseteq \partial \mathcal{S}_\mathsf{A} \cap \mathcal{S}_\mathsf{A}$ there exist a $p \in \mathbb{R}[x]$ with $p \geq 0$ and $L_{t''}(p) = 0$. By Lemma 5.2.5 we have $\operatorname{supp} \nu \subseteq \mathcal{Z}(p)$, i.e., ν has at most $\frac{d}{2}$ atoms. In summary, with

$$s_\mathsf{B}(1,\varepsilon) = \varepsilon^d \cdot s_\mathsf{B}((\varepsilon^{-1},1)) = \varepsilon^d \cdot s_\mathsf{A}(\varepsilon^{-1})$$

and the dehomogenisation we find that s has a $\frac{d}{2}+1$-atomic representing measure, i.e.,

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

It remains to show that $\mathcal{C}_\mathsf{A} \not\leq \frac{d}{2}$. Assume we have $\mathcal{C}_\mathsf{A} \leq \frac{d}{2}$. Then take $s \in \mathcal{S}_\mathsf{A}$ and a $\mathcal{C}_\mathsf{A}$ atomic representing measure μ with atoms at $x_1, \dots, x_{\mathcal{C}_\mathsf{A}}$. Then $p(x) := \prod_{i=1}^{\mathcal{C}_\mathsf{A}} (x - x_i)^2 \in \operatorname{Pos}(1,d)$ with $L_s(p) = 0$, i.e., $s \in \partial \mathcal{S}_\mathsf{A}$. Hence, $\mathcal{S}_\mathsf{A}$ has no interior which contradicts $\dim \mathcal{S}_\mathsf{A} = \dim \mathbb{R}[x]_{\leq d} = d+1$. Therefore, we have $\mathcal{C}_\mathsf{A} = \frac{d}{2} + 1$.

(b) Let d be odd: Since d is odd we have that $d+1$ is even. So let $s \in \mathcal{S}_\mathsf{A}$ be a moment sequence with a k-atomic representing measure μ. Since μ only contains atoms, we have that $s_{d+1} \int x^{d+1}\, \mathrm{d}\mu(x) \in [0,\infty)$, i.e., $s' := (s, s_{d+1}) \in \mathcal{S}_\mathsf{B}$ with $\mathsf{B} = \mathsf{A} \cup \{x^{d+1}\}$. But after homogenization we remove the atom $(1,0)$, to get a k-atomic representing measure with $k \leq \frac{d+1}{2}$ which represents s but not s' since s_{d+1} is altered by $s_\mathsf{B}((1,0)) = (0,\dots,0,1)$. Hence, after dehomogenization we find that s has a k-atomic representing measure with $k \leq \frac{d+1}{2} = \left\lceil \frac{d+1}{2} \right\rceil$. The last equality holds since d was odd. □

In the previous proof we actually proved in a side step the following homogeneous version.

Corollary 6.2.2. *Let $d \in \mathbb{N}$ and $\mathcal{V} = \mathbb{R}[x,y]_{=2d}$ with monomial basis B on $\mathcal{X} = \mathbb{P}^1 \cong S^1$. Then*

$$\mathcal{C}_\mathsf{B} = d+1$$

and $s \in \partial \mathcal{S}_\mathsf{B}$ if and only if $\mathcal{C}_\mathsf{B}(s) \leq d$.

6.3. A Lower Bound for differentiable Functions

Theorem 6.3.1. *Let $n \in \mathbb{N}$, $\mathcal{X} \subseteq \mathbb{R}^n$ open and $\mathcal{V} \subset C^{n\cdot d+1}(\mathcal{X},\mathbb{R})$ be a finite dimensional real vector space with basis A. Then*

$$\mathcal{C}_\mathsf{A} \geq \left\lceil \frac{\dim \mathcal{V}}{n+1} \right\rceil.$$

Proof. Let $d = \dim \mathcal{V}$, then $\mathcal{S}_\mathsf{A} \subseteq \mathbb{R}^d$ is full dimensional. From the moment map

$$S_{\mathsf{A},k} : [0,\infty)^k \times \mathcal{X}^k \to \mathbb{R}^d,\ (c_1,\dots,c_k,x_1,\dots,x_k) \mapsto \sum_{i=1}^k c_i \cdot s_\mathsf{A}(x_i)$$

and therefore the total derivative

$$\begin{aligned} &DS_{\mathsf{A},k}(c,x)\\ &= \Big(s_\mathsf{A}(x_1),\dots,s_\mathsf{A}(x_k), c_1\cdot \partial_1 s_\mathsf{A}(x_1),\dots,c_1\cdot\partial_n s_\mathsf{A}(x_k), c_2\cdot s_\mathsf{A}(x_2),\dots,c_k\cdot s_\mathsf{A}(x_k)\Big)\\ &\quad \in \mathbb{R}^{k(n+1)\times d}. \end{aligned}$$

A regular point (c,x) is such that $DS_{\mathsf{A},k}(c,x)$ has full rank, i.e., we need $k \geq \left\lceil \frac{d}{n+1} \right\rceil$. By linear independence of A we have that $k \leq d$. Hence, the regularity fulfills $n\cdot d+1 > n\cdot d = (n+1)\cdot d - d \geq \max\{0, k\cdot(n+1)-d\}$ and we can apply Sard's Theorem A.5.2. By Sard's Theorem A.5.2 the regular values of $S_{\mathsf{A},k}$ are dense in $\mathbb{R}^d$ and therefore dense in $\mathcal{S}_\mathsf{A}$, i.e., there are $s \in \mathcal{S}_\mathsf{A}$ which need at least $k \geq \left\lceil \frac{d}{n+1} \right\rceil$ atoms in a representing measure. □

Remark 6.3.2. H. M. Möller[22] [Möl76] gave for $\mathcal{V} = \mathbb{R}[x,y]_{\leq 2k-1}$, $k \in \mathbb{N}$, the lower bound

$$\mathcal{C}_{\mathsf{A}} \geq \binom{k+1}{2} + \left\lfloor \frac{k}{2} \right\rfloor =: M(k).$$

But from Theorem 6.3.1 we have the lower bound $\lceil \frac{1}{3}\binom{2k+1}{2}\rceil$ and hence for $k \geq 4$ we have

$$\left\lceil \frac{1}{3}\binom{2k+1}{2} \right\rceil - M(k) \geq \frac{(k-2)^2 - 4}{6}. \qquad \circ$$

Remark 6.3.3. For polynomials $\mathbb{R}[x_1,\dots,x_n]_{\leq d}$ the lower bound can be improved by the Alexander–Hirschowitz Theorem [AH95]. ○

Remark 6.3.4. The regularity $\mathcal{V} \subseteq C^{d\cdot n+1}(\mathcal{X},\mathbb{R})$ can be improved. But for $\mathcal{V} \subseteq C(\mathcal{X},\mathbb{R})$ Theorem 6.3.1 does in general not hold, e.g., for space filling curves. ○

Example 6.3.5. For any $d \in \mathbb{N}$ there exists a surjective $f = (f_1,\dots,f_d) \in C([0,1],[0,1]^d)$, i.e., a space filling curve [Sag94]. Let $\mathcal{V} = \operatorname{lin}\{f_1,\dots,f_d\}$ and therefore $\mathsf{A} = \{f_1,\dots,f_d\}$ be a basis of $\mathcal{V}$. Then $\mathcal{S}_{\mathsf{A}} = [0,\infty)^d$ and $\mathcal{C}_{\mathsf{A}} = 1$.

Proof. Since $f : [0,1] \to [0,1]^d$ is surjective we have $\mathcal{S}_{\mathsf{A}} = [0,\infty)^d$ and for any $s \in \mathcal{S}_{\mathsf{A}}$ we find a $c \in (0,1)$ such that $c \cdot s \in [0,1]^d$. Since f is surjective there exists a $x \in [0,1]$ such that $f(x) = c \cdot s$ and hence $s = \int_0^1 f(y)\, \mathrm{d}(c^{-1}\cdot\delta_x)(y)$. □

Remark 6.3.6. The previous example can even be extended to an infinite dimensional space by taking the $\aleph_0$-dimensional Schönberg[23] space filling curve $f : [0,1] \to [0,1]^{\aleph_0}$. ○

6.4. Lower Bounds on $\mathcal{C}(L)$ for $L : \mathbb{R}[x,y]_{\leq 2d} \to \mathbb{R}$ on $\mathcal{X} \subseteq \mathbb{R}^2$ open

Theorem 6.4.1. *Let $d \in \mathbb{N}_0$, $\mathcal{V} = \mathbb{R}[x,y]_{\leq 2d}$ with monomial basis A on $\mathcal{X} \subseteq \mathbb{R}^2$ with non-empty interior. Then*

$$\mathcal{C}_{\mathsf{A}} \geq d^2.$$

Proof. Since $\mathcal{X} \subseteq \mathbb{R}^2$ has non-empty interior we can (after a translation and scaling) assume that $\{1,2,\dots,d\}^2 \subseteq \mathcal{X}$. Under this translation and scaling $\mathcal{V} = \mathbb{R}[x,y]_{\leq 2d}$ remains unchanged.

Set $f(x,y) = (x-1)^2\cdots(x-d)^2 + (y-1)^2\cdots(y-d)^2 \in \operatorname{Pos}(2,2d)$ with $\mathcal{Z}(f) = \{1,\dots,d\}^2$. We want to apply Lemma 6.1.4, i.e., we show $\operatorname{rank}(s_{\mathsf{A}}(x,y))_{x,y=1,\dots,d} = d^2$. Since the row rank is equal to the column rank we have that

$$\operatorname{rank}(s_{\mathsf{A}}(x,y))_{x,y=1,\dots,d} = \dim \mathbb{R}[x,y]_{\leq 2d}\big|_{\{1,\dots,d\}^2}.$$

It is therefore sufficient to show that $\mathcal{W} := \mathbb{R}[x,y]_{\leq 2d}\big|_{\{1,\dots,d\}^2}$ has full dimension d^2. For that it is sufficient to show that any $p : \{1,\dots,d\}^2 \to \mathbb{R}$ is in $\mathcal{W}$. By linearity of $\mathcal{W}$ it

[22] H. Micheal Möller, Prof. i. R. TU Darmstadt
[23] Isaac Jacob Schoenberg (21 April 1903, Galati, Kingdom of Romania – 21 February 1990, ???)

is sufficient to show that for any $a, b = 1, \dots, d$ the characteristic function

$$\chi_{(a,b)}(x,y) := \begin{cases} 1 & \text{for } (x,y) = (a,b), \\ 0 & \text{otherwise} \end{cases}$$

is in $\mathcal{W}$. We have that

$$p_{(a,b)}(x,y) := c_{a,b} \cdot (x-1) \cdots \widehat{(x-a)} \cdots (x-d) \cdot (y-1) \cdots \widehat{(y-b)} \cdots (y-d)$$

for some $c_{a,b} \in \mathbb{R}$ is such a polynomial. $\square$

6.5. Lower Bounds on $\mathcal{C}(L)$ for $L : \mathbb{R}[x_1, \dots, x_n]_{\leq d} \to \mathbb{R}$ on $\mathcal{X} \subseteq \mathbb{R}^n$

Definition 6.5.1. Let $\mathbb{R}[x_0, \dots, x_n]$ be the polynomial ring with the natural grading and let $I \subset \mathbb{R}[x_0, \dots, x_n]$ be a homogeneous ideal. Let

$$R = \mathbb{R}[x_0, \dots, x_n]/I$$

be the quotient ring which is a graded ring itself. The *Hilbert function* HF_R of R is given by $HF_R(d) = \dim R_d$ where R_d is the degree d part of R.

Lemma 6.5.2. *Let $I \subset \mathbb{R}[x_1, \dots, x_n]$ be an ideal and $I^h \subset \mathbb{R}[x_0, \dots, x_n]$ be the* homogenization *of I, i.e., the ideal generated by the homogenizations f^h for all $f \in I$. Then the dehomogenization map induces an isomorphism of vector spaces*

$$(\mathbb{R}[x_0, \dots, x_n]/I^h)_d \quad \to \quad (\mathbb{R}[x_1, \dots, x_n]/I)_{\leq d}$$

for all $d \geq 0$.

Proof. Clear. $\square$

Example 6.5.3. Let $n \in \mathbb{N}$, $I = (0)$ and $R := \mathbb{R}[x_0, \dots, x_n]/I = \mathbb{R}[x_0, \dots, x_n]$. Then the Hilbert function HF_R is given by

$$HF_R(d) = HF_{\mathbb{R}[x_0,\dots,x_n]} = \binom{n+d}{n}. \qquad \circ$$

Lemma 6.5.4. *Let $n, d \in \mathbb{N}$ and A be the monomial basis of $\mathbb{R}[x_1, \dots, x_n]_{\leq d}$. Let $\Gamma \subset \mathbb{R}^n$ be a set of finitely many points and I be its vanishing ideal. Then*

$$\dim \operatorname{lin} \{s_{\mathsf{A}}(x) \,|\, x \in \Gamma\} = \dim(\mathbb{R}[x_1, \dots, x_n]/I)_{\leq d} = \dim(\mathbb{R}[x_0, \dots, x_n]/I^h)_d = HF_I(d).$$

Proof. Follows from the definition of the Hilbert function and Lemma 6.5.2. $\square$

Definition 6.5.5. Let R be a commutative ring. A sequence $f_1, \dots, f_r \in R$ is a *regular sequence* if for all $i = 1, \dots, r$ the residue class of f_i is not a zero divisor in $R/(f_1, \dots, f_{i-1})$.

Lemma 6.5.6. *Let $I \subset \mathbb{R}[x_0,\dots,x_n]$ be a homogeneous ideal and $R = \mathbb{R}[x_0,\dots,x_n]/I$ with Hilbert function HF_R. Let $f_1,\dots,f_r \in R$ be a regular sequence of homogeneous elements of degree d. Then the Hilbert function $HF_{R/(f_1,\dots,f_r)}$ of $R/(f_1,\dots,f_r)$ is*

$$HF_{R/(f_1,\dots,f_r)}(j) = \sum_{i=0}^{r} (-1)^i \cdot \binom{r}{i} \cdot HF_R(j - i\cdot d).$$

Proof. We prove the statement by induction on r. The case $r = 0$ is clear. In order to prove the induction step, let $R^i = R/(f_1,\dots,f_i)$ for $i = 0,\dots,r$. For all $j \in \mathbb{Z}$ we have the exact sequence

$$0 \quad \to \quad R^{r-1}_{j-d} \quad \xrightarrow{\cdot f_r} \quad R^{r-1}_j \quad \to \quad R^r_j \quad \to \quad 0.$$

Therefore,

$$HF_{R^r}(j) = HF_{R^{r-1}}(j) - HF_{R^{r-1}}(j-d).$$

By induction hypothesis this implies that

$$\begin{aligned} HF_{R^r}(j) &= \sum_{i=0}^{r-1} (-1)^i \cdot \binom{r-1}{i} \cdot HF_R(j - i\cdot d)\\ &\quad - \sum_{i=0}^{r-1} (-1)^i \cdot \binom{r-1}{i} \cdot HF_R(j - (i+1)\cdot d)\\ &= \sum_{i=0}^{r} (-1)^i \cdot \left[\binom{r-1}{i} + \binom{r-1}{i-1}\right] \cdot HF_R(j - i\cdot d)\\ &= \sum_{i=0}^{r} (-1)^i \cdot \binom{r}{i} \cdot HF_R(j - i\cdot d). \end{aligned}$$

□

Lemma 6.5.7. *Let $n, d \in \mathbb{N}$ and set*

$$p_i := (x_i - x_0)\cdots(x_i - dx_0)$$

for $i = 1,\dots,n$. The following holds:

(i) The sequence $p_1,\dots,p_n$ is regular.

(ii) The ideal generated by $p_1,\dots,p_n$ is radical.

(iii) Let $f_1,\dots,f_n$ be a regular sequence of homogeneous functions f_i of degree d. The Hilbert function HF_{R_n} of $R_n := \mathbb{R}[x_0,\dots,x_n]/(f_1,\dots,f_n)$ is

$$HF_{R_n}(k) = \sum_{i=0}^{n} (-1)^i \cdot \binom{n}{i} \cdot HF_{R_0}(k - i\cdot d).$$

In particular, we have

$$HF_{R_n}(2d) = \binom{n+2d}{n} - n\cdot\binom{n+d}{n} + \binom{n}{2}$$

and

$$HF_{R_n}(2d+1) = \binom{n+2d+1}{n} - n\cdot\binom{n+d+1}{n} + 3\cdot\binom{n+1}{3}.$$

Proof. (i): p_i is a monic polynomial over $\mathbb{R}[x_0]$ in the single variable x_i.

(ii): Follows from [Alo99, Thm. 1.1].

(iii): Since $HF_{R_0}(k) = \binom{n+k}{n}$ for $k \geq 0$ and $HF_{R_0}(k) = 0$ otherwise, Lemma 6.5.6 implies the statement. $\square$

Hilbert's Nullstellensatz 6.5.8. *Let $\mathbb{K}$ be an algebraically closed field and $J \subseteq \mathbb{K}[x_1,\dots,x_n]$ be an ideal. Then*

$$I(V(J)) = \sqrt{J}$$

where we have

$$\begin{aligned} V(J) &:= \{x\in\mathbb{K}^n \,|\, f(x)=0 \text{ for all } f\in J\},\\ I(V) &:= \{f\in\mathbb{K}[x_1,\dots,x_n] \,|\, f(x)=0 \text{ for all } x\in V\}, \text{ and}\\ \sqrt{J} &:= \{f\in\mathbb{K}[x_1,\dots,x_n] \,|\, f^r\in J \text{ for some } r\in\mathbb{N}\}. \end{aligned}$$

Theorem 6.5.9. *Let $n,d\in\mathbb{N}$ and $\mathcal{X}\subseteq\mathbb{R}^n$ with non-empty interior. For even degree $\mathcal{V} = \mathbb{R}[x_1,\dots,x_n]_{\leq 2d}$ we have*

$$\mathcal{C}_{\mathsf{A}_{n,2d}} \geq \binom{n+2d}{n} - n\cdot\binom{n+d}{n} + \binom{n}{2}$$

and for odd degree $\mathcal{V} = \mathbb{R}[x_1,\dots,x_n]_{\leq 2d+1}$ we have

$$\mathcal{C}_{\mathsf{A}_{n,2d+1}} \geq \binom{n+2d+1}{n} - n\cdot\binom{n+d+1}{n} + 3\cdot\binom{n+1}{3}.$$

Proof. Since $\mathcal{X}\subseteq\mathbb{R}^n$ has non-empty interior, there is a $\varepsilon>0$ and $y\in\mathbb{R}^n$ such that $y+\varepsilon\cdot\{1,\dots,d\}^n\subset\mathcal{X}$. The affine map $T:\mathcal{X}'\to\mathcal{X}$, $x\mapsto y+\varepsilon\cdot x$ shifts the moment problem on $\mathcal{X}$ to $\mathcal{X}' = \varepsilon^{-1}\cdot(\mathcal{X}-y)$ with $\mathbb{R}[x_1,\dots,x_d]_{\leq D} = \mathbb{R}[x_1,\dots,x_n]_{\leq D}\circ T$ with $D = 2d$ or $2d+1$ and hence $\{1,\dots,d\}^n\subset\mathcal{X}'$. So w.l.o.g. we can assume $\Gamma := \{1,\dots,d\}^n\subset\mathcal{X}$.

Set $p_i = (x_i-1)\cdots(x_i-d)$ and p_i^h their homogenizations. Then by Lemma 6.5.7(i) the sequence $p_1^h,\dots,p_n^h$ is regular and $I := (p_1^h,\dots,p_n^h)$ is radical, i.e., $V(I)=\Gamma^h$ and by Hilbert's Nullstellensatz 6.5.8 we have $I(\Gamma^h) = I(V(I)) = \sqrt{I} = I$. Then

$$\mathcal{C}_{\mathsf{A}} \overset{\text{Lemma 6.1.4}}{\geq} \dim\operatorname{lin}\{s_{\mathsf{A}}(x) \,|\, x\in\Gamma\} \overset{\text{Lemma 6.5.4}}{=} HF_I(D).$$

The $HF_I(D)$ are then given in Lemma 6.5.7(iii). $\square$

Example 6.5.10. Let $D\in\mathbb{N}$, $d := \lfloor\frac{D}{2}\rfloor$, and let $L:\mathbb{R}[x_1,\dots,x_n]_{\leq D}\to\mathbb{R}$ be a moment functional with representing

$$\mu = \sum_{x\in\{1,\dots,d\}^n} \delta_x.$$

Then

$$\mathcal{C}(L) = \begin{cases} \binom{n+2d}{n} - n\cdot\binom{n+d}{n} + \binom{n}{2} & \text{for } D = 2d,\\[2ex] \binom{n+2d+1}{n} - n\cdot\binom{n+d+1}{n} + 3\cdot\binom{n+1}{3} & \text{for } D = 2d+1.\end{cases} \qquad \circ$$

Corollary 6.5.11. *Let $n,d\in\mathbb{N}$ and $\mathcal{X}\subseteq\mathbb{R}^n$ with non-empty interior. Let $\mathcal{V}_{n,d} = \mathbb{R}[x_1,\dots,x_n]_{\leq d}$ with monomial basis $\mathsf{A}_{n,d}$. The following hold:*

(i) $\liminf_{d\to\infty} \frac{\mathcal{C}_{\mathsf{A}_{n,d}}}{\#\mathsf{A}_{n,d}} \geq 1 - \frac{n}{2^n}$ *for all $n\in\mathbb{N}$.*

(ii) For any $d\geq 4$ and $\varepsilon>0$ there exists a $n\in\mathbb{N}$ large enough such that there exists a moment functional $L:\mathbb{R}[x_1,\dots,x_n]_{\leq d}\to\mathbb{R}$ on $\mathcal{X}$ with

$$\mathcal{C}_{\mathsf{A}_{n,d}}(L) \quad\geq\quad (1-\varepsilon)\cdot\binom{n+d}{n}.$$

6.6. $\{s_\mathsf{A}(x)\,|\,x\in\mathcal{X}\}$ Countable

Theorem 6.6.1. *Let $\mathcal{X}$ be a set and $\mathcal{V}$ be a finite-dimensional vector space of real functions $f:\mathcal{X}\to\mathbb{R}$ with basis A and let $\{s_\mathsf{A}(x)\,|\,x\in\mathcal{X}\}$ be countable. Then*

$$\mathcal{C}_\mathsf{A} = \dim\mathcal{V}.$$

Proof. Let $P=\{p_1,\dots,p_k\}\subseteq\mathcal{X}$ with $k=|P|<\dim\mathcal{V}$. Then the cone C_P spanned by $s_\mathsf{A}(p_1),\dots,s_\mathsf{A}(p_k)$ is at most $\dim\mathcal{V}-1$ dimensional. Since $\mathcal{X}$ is countable and $|P|<\dim\mathcal{V}$, there are only countably many such $P\subseteq\mathcal{X}$. But since $\mathcal{S}_\mathsf{A}$ is full-dimensional, it is not the countable union of cones of dimension at most $\dim\mathcal{V}-1$. Hence, there are moment sequences which need $\dim\mathcal{V}$ point evaluations. $\square$

Part II. Applications and More

7. Finding atomic representing Measures

7.1. Flat Extension

Definition 7.1.1. Let $d, D \in \mathbb{N}_0$ be such that $d < D$. Let $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\leq 2D}$ be a real sequence and set $s^0 = (s_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\leq 2d}$. We say s is *flat* with respect to s^0 if

$$\operatorname{rank} \mathcal{H}_D(s) = \operatorname{rank} \mathcal{H}_d(s^0).$$

We say a linear functional $L : \mathbb{R}[x_1, \dots, x_n]_{\leq 2D} \to \mathbb{R}$ is *flat* if the corresponding sequence s with $s_\alpha = L(x^\alpha)$ for all α with $|\alpha| \leq 2D$ is flat, i.e., there is a $d < D$ such that s is flat with respect to $s^0 = (s_\alpha)_{\alpha:|\alpha|\leq 2d}$.

Flat Extension Theorem 7.1.2. *Let $d, D \in \mathbb{N}_0$ with $d < D$. If $L : \mathbb{R}[x_1, \dots, x_n]_{\leq 2D} \to \mathbb{R}$ is flat with respect to $L_0 = L|_{\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}}$ then L has a unique extension to a linear functional $\tilde{L} : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ such that $\tilde{L}$ is flat with respect to L_0. If $L(p^2) \geq 0$ for all $p \in \mathbb{R}[x_1, \dots, x_n]_{\leq 2d}$, then $\tilde{L}(q^2) \geq 0$ for all $q \in \mathbb{R}[x_1, \dots, x_n]$.*

The unique extension of a flat functional is called *flat extension*.

The proof of the theorem is too lengthy for the lecture. We therefore refer the reader to the original literature of Curto[24] and Fialkow[25] [CF96, CF98]. See also [Lau09] and [Sch17]. The main application is the following.

Theorem 7.1.3. *Let $d, D \in \mathbb{N}$ and $L : \mathbb{R}[x_1, \dots, x_n]_{\leq 2D} \to \mathbb{R}$ be a flat linear functional with respect to $L_0 := L|_{\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}}$ with $L(p^2) = L_0(p^2) \geq 0$ for all $p \in \mathbb{R}[x_1, \dots, x_n]_{\leq 2d}$. Then L is a moment functional with a $\operatorname{rank} \mathcal{H}(L)$-atomic representing measure.*

Proof. Since L is flat with respect to L_0 and $L_0(p^2) \geq 0$ for all $p \in \mathbb{R}[x_1, \dots, x_n]_{\leq d}$ we have by the Flat Extension Theorem 7.1.2 that there exists a unique functional $\tilde{L} : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ such that $\tilde{L}$ is flat with respect to L_0 and $\tilde{L}(p^2) \geq 0$ for all $p \in \mathbb{R}[x_1, \dots, x_n]$.

Since $\tilde{L}(p^2) \geq 0$ for all $p \in \mathbb{R}[x_1, \dots, x_n]$ we have

$$\tilde{L}(fg) \leq \tilde{L}(f^2) \cdot \tilde{L}(g^2)$$

for all $f, g \in \mathbb{R}[x_1, \dots, x_n]$. Then

$$\mathcal{N}_{\tilde{L}} := \{f \in \mathbb{R}[x_1, \dots, x_n] \,|\, \tilde{L}(fg) = 0 \text{ for all } g \in \mathbb{R}[x_1, \dots, x_n]\}$$

[24] Raúl Enrique Curto (unknown)
[25] Lawrance Fialkow (unknown)

is a radical[26] ideal in $\mathbb{R}[x_1,\dots,x_n]$. Then

$$\mathcal{D}_{\tilde{L}} := \mathbb{R}[x_1,\dots,x_n]/\mathcal{N}_{\tilde{L}}$$

is finite dimensional with $\dim \mathcal{D}_{\tilde{L}} = \operatorname{rank}\mathcal{H}(L) =: r$. To see this let $h^{(i)} = (h_{\alpha^{(i)},\beta})_{\beta\in\mathbb{N}_0^n}$ be column vectors of $H(\tilde{L})$. For $p(x) = \sum_i c_i x^{\alpha^{(i)}}$ we have

$$\sum_{i=1}^r c_i h_{\alpha^{(i)},\beta} = \tilde{L}\left(\sum_{i=1}^r c_i x^{\alpha^{(i)}+\beta}\right) = \tilde{L}(p(x)\cdot x^\beta)$$

for all $\beta\in\mathbb{N}_0^n$. Hence,

$$h^{(1)},\dots,h^{(m)} \text{ linearly independent} \quad\Leftrightarrow\quad x^{\alpha^{(1)}},\dots,x^{\alpha^{(m)}} \text{ linearly independent in } \mathcal{D}_{\tilde{L}}$$

and therefore $r = \operatorname{rank}\mathcal{H}(\tilde{L}) = \dim\mathcal{D}_{\tilde{L}}$. Hence,

$$(\mathcal{D}_{\tilde{L}}, \langle\,\cdot\,,\,\cdot\,\rangle) \quad \text{with} \quad \langle f,g\rangle := \tilde{L}(f,g)$$

is a finite dimensional (real) Hilbert space. For $i=1,\dots,n$ define the multiplication operators

$$M_i : \mathcal{D}_{\tilde{L}} \to \mathcal{D}_{\tilde{L}}, \quad f\mapsto x_i\cdot f,$$

then $(M_1,\dots,M_n)$ is a set of commuting symmetric operators on a finite dimensional Hilbert space and hence they possess a common set of (real) eigenvectors $e_1,\dots,e_r$ and have the eigenvalues $y_{i,j}$:

$$M_i e_j = y_{i,j} e_j$$

for all $i=1,\dots,n$ and $j=1,\dots,r$. Set $y_j := (y_{1,j},\dots,y_{n,j})^T\in\mathbb{R}^n$ for $j=1,\dots,r$. Since $1\in\mathcal{D}_{\tilde{L}}$ and $e_1,\dots,e_r$ a basis of $\mathcal{D}_{\tilde{L}}$, we write $1=\sum_{i=1}^r m_i e_i$ and we set $\mu := \sum_{i=1}^r m_i^2\delta_{y_i}$. Then

$$\begin{aligned}\tilde{L}(p) &= \langle p(x)\cdot 1, 1\rangle\\ &= \sum_{i,j=1}^r \langle p(M_1,\dots,M_n) m_i e_i, m_j e_j\rangle\\ &= \sum_{i=1}^r p(y_j) m_i^2 = \int p(y)\ \mathrm{d}\mu(y).\end{aligned}$$

□

Lemma 7.1.4. *Let $n,d\in\mathbb{N}$, $\mathsf{A}_{n,d}$ be the monomial basis of $\mathbb{R}[x_1,\dots,x_n]_{\leq d}$ and resp. $\mathsf{A}_{n,2d}$ of $\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}$. Let $y_1,\dots,y_k\in\mathbb{R}^n$ for some $k\in\mathbb{N}$. Let $L:\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}\to\mathbb{R}$ be represented by $\mu=\sum_{i=1}^k c_i\delta_{y_i}$. Then*

$$\mathcal{H}_d(L) = \sum_{i=1}^k c_i\cdot s_{\mathsf{A}_{n,d}}(y_i)\cdot s_{\mathsf{A}_{n,d}}(y_i)^T. \tag{5}$$

[26] Assume $1\notin\mathcal{N}_{\tilde{L}}$, otherwise $\mathcal{N}_{\tilde{L}} = \mathbb{R}[x_1,\dots,x_n]$ and $\tilde{L}=0$. If $f^k\in\mathcal{N}_{\tilde{L}}$ for some $k\geq 2$, then $|\tilde{L}(f^{k-1})|\leq\tilde{L}(1)\cdot\tilde{L}(f^{k+(k-2)})=0$.

and

$$\operatorname{rank}\mathcal{H}_d(L) \leq k.$$

The following are equivalent:

(i) $\operatorname{rank}\sum_{i=1}^k s_{\mathsf{A}_{n,d}}(y_i)\cdot s_{\mathsf{A}_{n,d}}(y_i)^T = k$.

(ii) $\delta_{y_1},\dots,\delta_{y_k}$ *are linearly independent over* $\mathbb{R}[x_1,\dots,x_n]_{\leq d}$.

(iii) $s_{\mathsf{A}_{n,d}}(y_1),\dots,s_{\mathsf{A}_{n,d}}(y_k)$ *are linearly independent.*

Proof. $s_x(y_i)\cdot s_{\mathsf{A}}(x)^T$ has the entries $x^{\alpha+\beta}$ and since the matrix $\mathcal{H}(L)$ is a sum of k matrices $s_{\mathsf{A}}(y_i)\cdot s_{\mathsf{A}}(y_i)^T$ and the rank of $s_{\mathsf{A}}(y_i)\cdot s_{\mathsf{A}}(y_i)^T$ is one, we have $\operatorname{rank}\mathcal{H}_d(L)\leq k$.

(i) $\Leftrightarrow$ (ii): Let $f=\sum_{\alpha:|\alpha|\leq d} f_\alpha x^\alpha$ and $\overrightarrow{f}=(f_\alpha)$ be the coefficient vector. Let $h^{(\alpha)}$ be the α-th column of $H_d(L)$ and e_α be the α-th basis vector. Then

$$\mathcal{H}_d(L)\overrightarrow{f} = \sum_\alpha f_\alpha \mathcal{H}_d(L)e_\alpha = \sum_\alpha f_\alpha h^{(\alpha)} = \sum_\alpha f_\alpha \sum_{j=1}^k c_j y_j^\alpha s_{\mathsf{A}_{n,d}}(y_j) = \sum_{j=1}^k c_j l_{x_j}(f) s_{\mathsf{A}_{n,d}}(y_j).$$

Since $c_i\neq 0$ and $\operatorname{range}\mathcal{H}_d(L)$ is contained in the span of vectors $s_{\mathsf{A}_{n,d}}(y_1),\dots,s_{\mathsf{A}_{n,d}}(y_k)$, it follows $\operatorname{rank}\mathcal{H}_d(L)=\dim\operatorname{range}\mathcal{H}_d(L)=k$ if and only if the $\delta_{y_1},\dots,\delta_{y_k}$ are linearly independent on $\mathbb{R}[x_1,\dots,x_n]_{\leq d}$.

(ii) $\Leftrightarrow$ (iii): Clear. $\square$

Theorem 7.1.5. *For every* $d\geq 2$ *there is a* $N\in\mathbb{N}$ *such that for every* $n\geq N$ *there exists a moment functional* $L:\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}\to\mathbb{R}$ *which has no extension to* $L':\mathbb{R}[x_1,\dots,x_n]_{\leq 4d}\to\mathbb{R}$ *with* L' *flat but* $L'':\mathbb{R}[x_1,\dots,x_n]_{\leq 4d+2}\to\mathbb{R}$ *is flat.*

Proof. By the $(d-1)$-Theorem 6.1.6 we have $C:=\mathcal{C}(L)\leq\binom{n+2d}{n}-1$, i.e., set $\mu=\sum_{i=1}^C\delta_{x_i}$ with δ_{x_i} linearly independent on $\mathbb{R}[x_1\dots,x_n]_{\leq 2d}$. Then $L_\infty:\mathbb{R}[x_1,\dots,x_n]\to\mathbb{R}$ has finite rank ($\operatorname{rank}\mathcal{H}(L_\infty)=C$) and is therefore flat.

Let $L:\mathbb{R}[x_1,\dots,x_n]_{\leq 2d}\to\mathbb{R}$ as in Example 6.5.10. Let $D\in\mathbb{N}$ be the smallest such that the extension L_{D+1} to $\mathbb{R}[x_1,\dots,x_n]_{\leq 2D+2}$ is flat but $L_\infty|_{\mathbb{R}[x_1,\dots,x_n]_{\leq 2D}}$ is not. Assume $D=2d-c$ for some $c\in\mathbb{N}$. From the condition $\mathcal{C}(L)\leq\binom{n+D}{n}$ that the Hankel matrix of the flat extension must be at least the size of the Carathéodory number of L we find that

$$\begin{aligned}1\leq\lim_{n\to\infty}\frac{\binom{n+2d-c}{n}}{\mathcal{C}(L)} &= \lim_{n\to\infty}\frac{\binom{n+2d-c}{n}}{\binom{n+2d}{n}}\cdot\underbrace{\frac{\binom{n+2d}{n}}{\mathcal{C}(L)}}_{\to 1\text{ by Cor. 6.5.11}}\\ &= \lim_{n\to\infty}\frac{(2d-c+1)\cdots(2d)}{(n+2d-c+1)\cdot(n+2d)}=0.\end{aligned}$$

A contradiction, i.e., $c=0$ must hold and therefore $D=2d$. $\square$

Example 7.1.6. Consider the moment functional $L : \mathbb{R}[x_1,\dots,x_n]_{\leq 2d} \to \mathbb{R}$ from Example 6.5.10 supported on the grid $\{1,\dots,d\}^n$. For this example we have that

$$(n,d) = (9,2),\ (7,3),\ \text{and}\ (6,4)$$

are already small examples where we have to extend it to the worst cast $D = 2d$. Even for $d = 10^{15}$ the worst case extension is already necessary for $n = 51$. ○

7.2. The generalized Eigenvalue Problem and Finding Atomic Representing Measures for $L : \mathbb{R}[x_1,\dots,x_n]_{\leq 2d} \to \mathbb{R}$ on $\mathcal{X} = \mathbb{R}^n$

Definition 7.2.1. Let $m \in \mathbb{N}$ and $A, B \in \mathbb{R}^{m\times m}$ be two symmetric matrices. A vector $v \in \mathbb{R}^m$ is called an *generalized eigenvector* and $\lambda \in \mathbb{R}$ the corresponding *generalized eigenvalue* if $Av \neq 0$ and

$$\lambda A v = B v \tag{6}$$

holds. (6) is called *generalized eigenvalue problem* (*for A and B*).

Note, even for symmetric matrices $A \in \mathbb{R}^{m\times m}$, there are only $\operatorname{rank} A$ many eigenvectors and -values (counting multiplicities).

Example 7.2.2. Let

$$A = \begin{pmatrix} 5 & -5 & 11 & -17 & 35 \\ -5 & 11 & -17 & 35 & -65 \\ 11 & -17 & 35 & -65 & 131 \\ -17 & 35 & -65 & 131 & -257 \\ 35 & -65 & 131 & -257 & 515 \end{pmatrix}$$

and

$$B = \begin{pmatrix} -5 & 11 & -17 & 35 & -65 \\ 11 & -17 & 35 & -65 & 131 \\ -17 & 35 & -65 & 131 & -257 \\ 35 & -65 & 131 & -257 & 515 \\ -65 & 131 & -257 & 515 & -1025 \end{pmatrix}.$$

Then the generalized eigenvalue problem $\lambda A v = B v$ has the eigenvalues $\lambda = -2, -1, 1$. ○

Theorem 7.2.3. *Let $L : \mathbb{R}[x_1,\dots,x_n]_{\leq 2d}$ be a flat moment functional and let $\mu = \sum_{i=1}^k c_i \delta_{y_i}$, $k = \mathcal{C}(L)$, be the unique representing measure with $c_i > 0$ and $y_i = (y_{i,1},\dots,y_{i,n}) \in \mathbb{R}^n$. Then for each $j = 1,\dots,n$ we have that the j-th coordinates $y_{1,j},\dots,y_{k,j}$ are exactly the set of generalized eigenvalues (counting multiplicities) of*

$$\lambda \mathcal{H}_{d-1}(L) v = \mathcal{H}_{d-1}(X_j L) v.$$

In fact there are $v_1, \dots, v_k \in \operatorname{range} \mathcal{H}_{d-1}(L) \setminus \{0\}$ *such that* $\mathcal{H}_{d-1}(L)v_i \neq 0$ *and*

$$y_{i,j}\mathcal{H}_{d-1}(L)v_i = \mathcal{H}_{d-1}(X_j L)v_i$$

for all $i = 1, \dots, k$ *and* $j = 1, \dots, n$. *For all other* $v \in \operatorname{range} \mathcal{H}_{d-1}(L)^\perp$ *we have* $\mathcal{H}_{d-1}(L)v = 0$.

Proof. Since L is flat we have that by Lemma 7.1.4

$$\mathcal{H}_{d-1}(L) = \sum_{i=1}^{k} c_i \cdot s_{\mathsf{A}_{n,d-1}}(y_i) \cdot s_{\mathsf{A}_{n,d-1}}(y_i)^T$$

and

$$\mathcal{H}_{d-1}(X_j L) = \sum_{i=1}^{k} c_i \cdot y_{i,j} \cdot s_{\mathsf{A}_{n,d-1}}(y_i) \cdot s_{\mathsf{A}_{n,d-1}}(y_i)^T$$

as well as that $s_1 := s_{\mathsf{A}_{n,d-1}}(y_1), \dots, s_k := s_{\mathsf{A}_{n,d-1}}(y_k)$ are linearly independent.

Hence, for each $i = 1, \dots, k$ we therefore have that there exists a $v_i \in \operatorname{range} \mathcal{H}_{d-1}(L) \setminus \{0\}$ such that $\langle s_l, v_i \rangle = \delta_{l,i}$ for all $l = 1, \dots, k$. We therefore have

$$\mathcal{H}_{d-1}(L)v_i = \sum_{i'=1}^{k} c_{i'} \cdot s_{i'} \cdot \underbrace{s_{i'}^T \cdot v_i}_{=\delta_{i,i'}} = c_i \cdot s_i \neq 0$$

and

$$\mathcal{H}_{d-1}(X_j L)v_i = \sum_{i'=1}^{k} c_{i'} \cdot y_{i',j} \cdot s_{i'} \cdot \underbrace{s_{i'}^T \cdot v_i}_{=\delta_{i,i'}} = c_i \cdot y_{j,i} \cdot s_i = y_{j,i} \cdot \mathcal{H}_{d-1}(X_j L)v_i.$$

Since $v_1, \dots, v_k$ is a basis of $\operatorname{lin}\{s_1, \dots, s_k\} = \operatorname{range} \mathcal{H}_{d-1}(L)$ we have that for all

$$v \in \operatorname{lin}\{s_1, \dots, s_k\}^\perp = \operatorname{lin}\{v_1, \dots, v_k\}^\perp = \operatorname{range} \mathcal{H}_{d-1}(L)^\perp$$

we have $\mathcal{H}_{d-1}(L)v = 0$. □

Example 7.2.4. Let

$$s = (5, -5, 11, -17, 35, -65, 131, -257, 515, -1025) \in \mathbb{R}^{10}.$$

Then $\mathcal{H}_4(s) = A$ and $\mathcal{H}_4(Xs) = B$ in Example 7.2.2 and hence the generalized eigenvalues are $\lambda = -2, -1, 1$. From linear algebra we find

$$s = 2s_{\mathsf{A}}(-2) + 2s_{\mathsf{A}}(-1) + s_{\mathsf{A}}(1),$$

i.e., s has the representing measure $\mu = 2\delta_{-2} + 2\delta_{-1} + \delta_1$ and is therefore also a moment sequence. ◦

8. Waring[27] Decomposition

8.1. The Apolar Scalar Product and Powers of Linear Forms

Definition 8.1.1. Let $n, d \in \mathbb{N}$. We set

$$\mathsf{N}_{n,d} := \{\alpha \in \mathbb{N}_0^n \,|\, |\alpha| = \alpha_1 + \dots + \alpha_n = d\}$$

and

$$\mathcal{H}_{n,d} := \mathbb{R}[x_1, \dots, x_n]_d.$$

Definition 8.1.2. Let $n, d \in \mathbb{N}$. With

$$\alpha! = \alpha_1! \cdots \alpha_n! \quad \text{and} \quad \binom{d}{\alpha} := \frac{d!}{\alpha!} = \frac{d!}{\alpha_1! \cdots \alpha_n!}$$

for $\alpha \in \mathsf{N}_{n,d}$ let

$$p(x) = \sum_{\alpha \in \mathsf{N}_{n,d}} \binom{d}{\alpha} \cdot a_\alpha \cdot x^\alpha \qquad \text{and} \qquad q(x) = \sum_{\alpha \in \mathsf{N}_{n,d}} \binom{n}{\alpha} \cdot a_\alpha \cdot x^\alpha \tag{7}$$

be homogeneous polynomials in $\mathcal{H}_{n,d}$. We define the *apolar scalar product* $[\,\cdot\,,\,\cdot\,]$ by

$$[p, q] := \sum_{\alpha \in \mathsf{N}_{n,d}} \binom{d}{\alpha} \cdot a_\alpha \cdot b_\alpha.$$

Lemma 8.1.3. *Let $n, d \in \mathbb{N}$. Then $(\mathcal{H}_{n,d}, [\,\cdot\,,\,\cdot\,])$ is a Hilbert space of dimension $\binom{n+d-1}{n-1}$.*

Proof. Clear. □

Corollary 8.1.4. *Let $n, d \in \mathbb{N}$ and $p \in \mathcal{H}_{n,d}$ written as in (7). Then*

$$[p, x^\alpha] = a_\alpha$$

for all $\alpha \in \mathsf{N}_{n,d}$.

Definition 8.1.5. Let $n, d \in \mathbb{N}$ and $y = (y_1, \dots, y_n) \in \mathbb{R}^n$. We denote by $a \cdot b = a_1 b_1 + \dots + a_n b_n$ the standard scalar product for all $a, b \in \mathbb{R}^n$. We define

$$(y \cdot\,)^d(x) := (y \cdot x)^d = \sum_{\alpha \in \mathbb{N}_{n,d}} \binom{d}{\alpha} \cdot x^\alpha \cdot y^\alpha$$

the d-th power of the linear form $y \cdot x$.

---

[27] Edward Waring (1736, Old Heath (near Shrewsbury, UK) – 15 August 1798, Pontesbury (Shropshire, UK))

Theorem 8.1.6. *Let $n, d \in \mathbb{N}$, $p \in \mathcal{H}_{n,d}$, and $y \in \mathbb{R}^n$. Then*

$$[p, (y\cdot\,)^d] = p(y).$$

Proof. Write p as in (7), then from Definition 8.1.5 we get

$$[p, (y\cdot\,)^d] = \sum_{\alpha\in\mathsf{N}_{n,d}} \binom{d}{\alpha} \cdot a_\alpha \cdot y^\alpha = p(y). \qquad \square$$

Remark 8.1.7. For $a, b \in \mathbb{R}^n$ we have $[(a\cdot\,)^d, (b\cdot\,)^d] = (a\cdot b)^d$, i.e., $a \perp b$ in $\mathbb{R}^d$ iff $(a\cdot\,)^d \perp (b\cdot\,)^d$ in $\mathcal{H}_{n,d}$. Hence, if $y_1, \dots, y_r \in \mathbb{R}^d$ are pairwise othogonal, then $(y_1\cdot\,)^d, \dots, (y_r\cdot\,)^d$ are pairwise orthogonal. ∘

Corollary 8.1.8. *Let $n, d \in \mathbb{N}$ and $U \subseteq \mathbb{R}^n$ be open and non-empty. Then*

$$\operatorname{lin}\{(y\cdot\,)^d \,|\, y \in U\} = \mathcal{H}_{n,d}.$$

Proof. $\subseteq$ is clear. To prove $\supseteq$ note that $(\mathcal{H}_{n,d}, [\,\cdot\,,\,\cdot\,])$ is a Hilbert space and hence let $p \in \operatorname{lin}\{(y\cdot\,)^d \,|\, y \in U\}^\perp \subseteq \mathcal{H}_{n,d}$. Then

$$p(y) \overset{\text{Thm. 8.1.6}}{=} [p, (y\cdot\,)^d] = 0$$

for all $y \in U$. Since p is a polynomial equal to zero on an open set $U \subseteq \mathbb{R}^n$ we have $p = 0$ which proves the inclusion $\supseteq$. $\square$

Theorem 8.1.9. *Let $n, d \in \mathbb{N}$. Then $\{(\alpha\cdot\,)^d \,|\, \alpha \in \mathsf{N}_{n,d}\}$ is a basis of $\mathcal{H}_{n,d}$.*

Proof. For $\beta \in \mathsf{N}_{n,d}$ we define

$$f_\beta(x) = \prod_{j=1}^{n} \prod_{i=0}^{\beta_j - 1} (dx_j - i(x_1 + \dots x_n)) \quad \in \mathcal{H}_{n,d}.$$

We have

$$f_\beta(\beta) = \prod_{j=1}^{n} \prod_{i=0}^{\beta_j - 1} (d\beta_j - id) = d^d \cdot \beta! \neq 0.$$

For $\alpha \neq \beta$ there is an index j such that $\alpha_j < \beta_j$ and hence an i with $i = \alpha_j$. Therefore, $f_\beta(\alpha)$ contains a factor $d\alpha_j - \alpha_j d = 0$, i.e., $f_\beta(\alpha) = 0$. In summary we have $[f_\beta, (\alpha\cdot\,)^d] = 0$ for $\alpha \neq \beta$ and $[f_\beta, (\beta\cdot\,)^d] \neq 0$ and therefore all $(\alpha\cdot\,)^d$ with $\alpha \in \mathsf{N}_{n,d}$ are linearly independent. Since $|\mathsf{N}_{n,d}| = \dim \mathcal{H}_{n,d}$ it is a basis. $\square$

Definition 8.1.10. Let $n, d \in \mathbb{N}$. $\{y_1, \dots, y_r\} \subset \mathbb{R}^n$ with $r = |\mathsf{N}_{n,d}|$ is called a *basic set of nodes* if $\{(y_1\cdot\,)^d, \dots, (y_r\cdot\,)^d\}$ is a basis of $\mathcal{H}_{n,d}$.

Definition 8.1.11. Let $n, d \in \mathbb{N}$. Let $F = \{f_\alpha \,|\, \alpha \in \mathsf{N}_{n,d}\}$ and $G = \{g_\alpha \,|\, \alpha \in \mathsf{N}_{n,d}\}$ be two bases of $\mathcal{H}_{n,d}$. We say that F and G are *dual bases* if

$$[f_\alpha, g_\beta] = \delta_{\alpha,\beta}$$

holds for all $\alpha, \beta \in \mathsf{N}_{n,d}$.

Remark 8.1.12. For dual bases F and G we have

$$h = \sum_{\alpha \in \mathsf{N}_{n,d}} [h, f_\alpha] \cdot g_\alpha = \sum_{\alpha \in \mathsf{N}_{n,d}} [h, g_\alpha] \cdot f_\alpha$$

for all $h \in \mathcal{H}_{n,d}$. ∘

Theorem 8.1.13. *Two subsets $F = \{f_\alpha \,|\, \alpha \in \mathsf{N}_{n,d}\}$ and $G = \{g_\alpha \,|\, \alpha \in \mathsf{N}_{n,d}\}$ of $\mathcal{H}_{n,d}$ for a set of dual bases if and only if they fulfill the* Marsden identity

$$(y \cdot x)^d = \sum_{\alpha \in \mathsf{N}_{n,d}} f_\alpha(y) \cdot g_\alpha(x)$$

for all $x, y \in \mathbb{R}^d$.

Proof. Suppose F and G are dual bases, then

$$(y\cdot\,)^d \overset{\text{Rem. 8.1.12}}{=} \sum_{\alpha \in \mathsf{N}_{n,d}} f_\alpha(y) \cdot g_\alpha \tag{8}$$

and hence the Marsden identity holds.

Conversely, assume that the Marsden identity holds. Then (8) holds on $\mathbb{R}^n$ and since by Corollary 8.1.8 the $(y\cdot\,)^d$ span $\mathcal{H}_{n,d}$ we have that G spans $\mathcal{H}_{n,d}$. But since $|G| = \dim \mathcal{H}_{n,d}$ we have that G is a basis. Then

$$g_\alpha(x) = [g_\alpha, (x\cdot\,)^d] = \sum_{\beta \in \mathsf{N}_{n,d}} [g_\alpha, f_\beta] \cdot g_\beta(x)$$

and since G is a basis we have $[g_\alpha, f_\beta] = \delta_{\alpha,\beta}$ for all $\alpha, \beta \in \mathsf{N}_{n,d}$. □

8.2. The Apolar Scalar Product and Differential Operators

Definition 8.2.1. Let $n, d \in \mathbb{N}$ and $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$. We define

$$\partial_i := \frac{\partial}{\partial x_i} \qquad \text{and} \qquad \partial^\alpha = \partial_1^{\alpha_1} \cdots \partial_n^{\alpha_n}.$$

In the same way we define $p(\partial)$ for $p \in \mathbb{R}[x_1, \dots, x_n]$.

Lemma 8.2.2. *Let $n, d, r \in \mathbb{N}$, $y_1, \dots, y_r \in \mathbb{R}^n$, $c_1, \dots, c_r \in \mathbb{R}$, and let*

$$f = \sum_{i=0}^{r} c_r \cdot (y_i\cdot\,)^d \quad \in \mathcal{H}_{n,d}.$$

For $p \in \mathcal{H}_{n,k}$ with $k \leq d$ we have

$$p(\partial) f = d \cdot (d-1) \cdots (d+1-k) \cdot \sum_{i=1}^{r} c_i \cdot p(y_i) \cdot (y_i\cdot\,)^{d-k}$$

Let $y \in \mathbb{R}^n$ and $p \in \mathcal{H}_{n,d}$, then

$$p(\partial)(y\cdot\,)^d = d! \cdot p(y).$$

In particular, if $p(y) = 0$, then $p(\partial)(y\cdot\,)^d = 0$.

Proof. Let $\alpha \in \mathsf{N}_{n,k}$. Then

$$\begin{aligned}\partial^\alpha (y\cdot x)^d &= \partial_1^{\alpha_1}\cdots\partial_n^{\alpha_n}(y\cdot x)^d\\ &= d\cdot(d-1)\cdots(d+1-k)\cdot y_1^{\alpha_1}\cdots y_n^{\alpha_n}\cdot (y\cdot x)^{d-k}\\ &= d\cdot(d-1)\cdots(d+1-k)\cdot y^\alpha\cdot(y\cdot x)^{d-k}.\end{aligned}$$

By linearity the statements follow. □

Lemma 8.2.3. *Let $n,d\in\mathbb{N}$ and $p,q\in\mathcal{H}_{n,d}$. Then*

$$[p,q] = \frac{1}{d!}p(\partial)q = \frac{1}{d!}q(\partial)p.$$

Proof. By symmetry of the apolar scalar product it is sufficient to prove the statement for $p(x) = \binom{d}{\alpha}x^\alpha$ and $q(x) = x^\beta$ for all $\alpha,\beta\in\mathsf{N}_{n,d}$.

$\underline{\alpha\neq\beta}$: Then $[p,q]=0$ and $\partial^\alpha x^\beta = 0$.

$\underline{\alpha=\beta}$: We have $\partial^\alpha x^\alpha = \alpha!$ and hence

$$\frac{1}{d!}p(\partial)q = \frac{1}{d!}\binom{d}{\alpha}\partial^\alpha x^\alpha = 1 = [\binom{d}{\alpha}x^\alpha, x^\alpha] = [p,q].$$ □

Corollary 8.2.4. *Let $f = \sum_{i=1}^k c_i\cdot(y_i\cdot\,)^d\in\mathcal{H}_{n,d}$ with $y_j\in\mathbb{R}^n$ and $c_i\in\mathbb{R}$, then*

$$\frac{1}{d!}f(\partial)p = [f,p] = \sum_{i=1}^k c_i\cdot p(y_j).$$

Lemma 8.2.5. *Let $n,d,D\in\mathbb{N}$ with $d<D$. For $p\in\mathcal{H}_{n,D}$, $q\in\mathcal{H}_{n,d}$, and $f\in\mathcal{H}_{n,D-d}$ we have*

$$D!\cdot[p,fq]_D = (D-d)!\cdot[f,q(\partial)p]_{D-d},$$

where $[\,\cdot\,,\,\cdot\,]_k$ is the apolar scalar product on $\mathcal{H}_{n,k}$.

Proof. With Lemma 8.2.3 we get

$$D!\cdot[p,fq]_D = (fq)(\partial)p = f(\partial)(q(\partial)p) = (D-d)!\cdot[f,q(\partial)p]_{D-d}.$$ □

8.3. Moment Functionals and Waring Decomposition

Definition 8.3.1. Let $n,d\in\mathbb{N}$. We set

$$\mathrm{War}(n,d) := \left\{ f\in\mathcal{H}_{n,d}\,\middle|\, f=\sum_{i=1}^k (y_i\cdot\,)^d \text{ for some } k\in\mathbb{N}_0,\ y_1,\dots,y_k\in\mathbb{R}^n\right\}$$

and

$$\mathrm{Pos}^h(n,2d) := \{f\in\mathcal{H}_{n,2d}\,|\, f\geq 0\}.$$

For $f\in\mathcal{W}(n,d)$ any $f=\sum_{i=1}^k (y_i\cdot\,)^d$ is called a *Waring decomposition of f*.

Corollary 8.3.2. *Let $n, d \in \mathbb{N}$. Then* $\mathrm{War}(n, 2d)$ *and* $\mathrm{Pos}^h(n, 2d)$ *are closed convex cones in* $\mathcal{H}_{n,2d}$ *with*

$$\mathrm{War}(n,2d) = \mathrm{Pos}^h(n,2d)^* \qquad \text{and} \qquad \mathrm{Pos}^h(n,2d) = \mathrm{War}(n,2d)^*.$$

Proof. That $\mathrm{War}(n, 2d)$ and $\mathrm{Pos}^h(n, 2d)$ are convex cones is clear. That $\mathrm{Pos}^h(n, 2d)$ is closed is also clear. We show $\mathrm{War}(n, 2d)$ is closed. Let $(f_k)_{k\in\mathbb{N}} \in \mathrm{War}(n, 2d)$ be a sequence such that $f_k \to f \in \mathcal{H}_{n,2d}$. By Carathéodory's Theorem A.4.1 we have

$$f_k = \sum_{i=1}^{C} (y_{k,i}\cdot\,)^{2d}$$

with $C \le \dim \mathcal{H}_{n,2d} < \infty$. Expanding the $2d$-th powers we find $(y_{k,i}\cdot x)^{2d} = (y_{k,i,1}x_1)^{2d} + \dots + (y_{k,i,n}x_n)^{2d} + \dots$ and since the coefficients converge we have that there is a $c > 0$ such that $\|y_{k,i}\| \le c$ for all $k \in \mathbb{N}$ and $i = 1, \dots, C$. Hence, by the Heine–Borel Theorem there is a subsequence $(k_j)_{j\in\mathbb{N}} \subseteq \mathbb{N}$ such that $y_{k_j,i} \to y_i \in \mathbb{R}^n$ for all $i = 1, \dots, C$ and hence $f = \sum_{i=1}^{C} (y_i\cdot\,)^{2d} \in \mathrm{War}(n, 2d)$ which proves closeness.

For the duality identities we have $0 \le f(y) = [f, (y\cdot\,)^{2d}]$ for all $y \in \mathbb{R}^n$ iff $0 \le [f, w]$ for all $w \in \mathrm{War}(n, 2d)$ which proves $\mathrm{Pos}^h(n, 2d) = \mathrm{War}(n, 2d)^*$. From $\mathrm{War}(n, 2d) \subseteq \mathrm{Pos}^h(n, 2d)^*$ we get

$$\mathrm{Pos}^h(n,2d) = \mathrm{War}(n,2d)^* \supseteq \mathrm{Pos}^h(n,2d)^{**} = \mathrm{Pos}^h(n,2d)$$

where the last equality holds by the bidual theorem. □

Theorem 8.3.3. *Let $n, d \in \mathbb{N}$.*

(i) For each $f \in \mathrm{War}(n, d)$ we have that $L_f : \mathcal{H}_{n,d} \to \mathbb{R}$ defined by

$$L_f(p) := \frac{1}{d!} f(\partial)p = [f, p]$$

for all $p \in \mathcal{H}_{n,d}$ is a moment functional.

(ii) For each moment functional $L : \mathcal{H}_{n,d} \to \mathbb{R}$ there is a unique $f \in \mathrm{War}(n, d)$ with $L = L_f$ and we have

$$\mu = \sum_{i=1}^{k} \delta_{y_i} \in \mathcal{M}(L) \qquad \Leftrightarrow \qquad f = \sum_{i=1}^{k} (y_i\cdot\,)^d.$$

Proof. (i): Since $f \in \mathrm{War}(n, d)$ we have $f = \sum_{i=1}^{k} (y_i\cdot\,)^d$ and hence $L_f(p) = \sum_{i=1}^{k} p(y_i) = \int p(x)\ \mathrm{d}\left(\sum_{i=1}^{k} \delta_{y_i}\right)(x)$.

(ii): Since $L : \mathcal{H}_{n,d} \to \mathbb{R}$ is a moment functional there exists by Richter's Theorem 5.3.1 a representing measure $\mu = \sum_{i=1}^{k} c_i \cdot \delta_{x_i}$ with $c_i > 0$. Hence $f = \sum_{i=1}^{k} c_i \cdot (x_i\cdot\,)^d = \sum_{i=1}^{k} (y_i\cdot\,)^d \in \mathrm{War}(n, d)$ with $y_i = \sqrt[d]{c_i} \cdot x_i$. Uniqueness follows from Riesz Representation Theorem. □

Theorem 8.3.4. *Let $n, d \in \mathbb{N}$ and set $h_{n,2d}(x) := \|x\|^{2d} = (x_1^2 + \cdots + x_n^2)^d$. Then*

$$h_{n,2d} \in \mathrm{War}(n, 2d)$$

and for any $u \in S^{n-1}$ there is a $\varepsilon > 0$ such that

$$h_{n,2d} = \varepsilon \cdot (u\cdot\,)^{2d} + \sum_{i=1}^{k} (y_i\cdot\,)^{2d}$$

for some $y_1, \dots, y_k \in \mathbb{R}^n$.

Proof. Denote by σ the surface measure on S^{n-1}. We define

$$h(x) := \int_{S^{n-1}} (y \cdot x)^{2d} \,\mathrm{d}\sigma(y)$$

for all $x \in \mathbb{R}^n$. Since $(y \cdot x)^{2d}$ is homogeneous in x we have $h \in \mathcal{H}_{n,2d}$. Since σ is invariant under rotation in $y \in \mathbb{R}^n$ we have that h is invariant under rotation in $x \in \mathbb{R}^n$ and therefore $h(x) = c \cdot \|x\|^{2d}$ with $c := h(\xi) > 0$ for any $\xi \in S^{n-1}$.

Define $L(f) := \int_{S^{n-1}} f(y) \,\mathrm{d}\sigma(y)$. Then $L(p) > 0$ for all $p \in \mathrm{Pos}^h(n, 2d) \setminus \{0\}$ and hence $L \in \operatorname{int} \mathrm{Pos}^h(n, 2d)^*$. Therefore, for any $u \in S^{n-1}$ there is a $\varepsilon = \varepsilon(u) > 0$ such that $\tilde{L} := L - \varepsilon\delta_u \in \mathrm{Pos}^h(n, 2d)^*$, i.e., $\tilde{L}$ is still a moment functional. By Richter's Theorem 5.3.1 $\tilde{L}$ has finitely atomic representing measure $\sum_{i=1}^k c_i\delta_{x_i}$ and hence L has a finitely atomic representing measure $\varepsilon\delta_u + \sum_{i=1}^k c_i\delta_{x_i}$. In summary,

$$c \cdot h_{n,2d} = L((y\cdot\,)^{2d}) = \varepsilon(u\cdot\,)^{2d} + \sum_{i=1}^{k} c_i(x_i\cdot\,)^{2d} = \sum_{i=0}^{k} (y_i\cdot\,)^{2d} \in \mathrm{War}(n, 2d). \qquad \square$$

We remind the reader that $\Delta := \partial_1^2 + \cdots + \partial_n^2$ is the Laplace operator.

Theorem 8.3.5. *Let $n, d \in \mathbb{N}$. Then*

$$L : \mathcal{H}_{n,2d} \to \mathbb{R}, \quad p \mapsto L(p) := \Delta^d p$$

is a moment functional such that $L(p) > 0$ for all $p \in \mathrm{Pos}^h(n, 2d) \setminus \{0\}$.

Proof. We have $\Delta^d = h_{n,2d}(\partial)$ and therefore $L(p) = h_{n,2d}(\partial)p = (2d)! \cdot [h_{n,2d}, p]$ by Theorem 8.3.3(i). Since $h_{n,2d} \in \mathrm{War}(n, 2d)$ by Theorem 8.3.4 we have that L is a moment functional by Theorem 8.3.3(i).

Let $p \in \mathrm{Pos}^h(n, 2d) \setminus \{0\}$. Then there is a $u \in S^{n-1}$ such that $p(u) > 0$ and therefore

$$L(p) = (2d)! \cdot [h_{n,2d}, p] \geq (2d)! \cdot [\varepsilon(u\cdot\,)^{2d}, p] = (2d)! \cdot \varepsilon \cdot p(u) > 0$$

by Theorem 8.3.4. $\square$

8.4. The Carathéodory Number and the Waring Rank

Hilbert's 17th problem states:

> For any $f \in \mathbb{R}[x_1,\dots,x_n]$, is it true that $f \geq 0$ on $\mathbb{R}^n$ implies that f is a sum of squares of rational functions $f = \sum_i \frac{p_i^2}{q_i^2}$?

Artin proved the general[28] case [Art27]. Pfister showed that 2^n squares are sufficient [Pfi67]. We are therefore also interested in how many d-powers are required in a Waring decomposition.

Definition 8.4.1. Let $n, d \in \mathbb{N}$. We define the *Waring rank $\mathcal{W}(f)$ of* $f \in \mathrm{War}(n, 2d)$ by

$$\mathcal{W}(f) := \min \left\{ k \in \mathbb{N}_0 \;\middle|\; f = \sum_{i=1}^{k} (y_i \cdot\,)^{2d} \text{ for some } y_1, \dots, y_k \in \mathbb{R}^n \right\}.$$

We define the *Waring rank* $\mathcal{W}(n, 2d)$ by

$$\mathcal{W}(n,2d) := \max_{f \in \mathrm{War}(n,2d)} \mathcal{W}(f).$$

Theorem 8.4.2. *Let $n, d \in \mathbb{N}$ and $f \in \mathrm{War}(n, 2d)$. Then*

$$\mathcal{W}(f) = \mathcal{C}(L_f).$$

Proof. Follows immediately from Theorem 8.3.3 (ii). □

Theorem 8.4.3. *Let $n, d \in \mathbb{N}$. Then*

$$\mathcal{W}(n,2d) \geq \binom{n+2d-1}{n-1} - (n-1) \cdot \binom{n+d-1}{n-1} + \binom{n-1}{2}.$$

Proof. Combine Theorem 8.3.3, Theorem 8.4.2, and use the special (affine) case from Example 6.5.10. Since Example 6.5.10 is affine we have by (de)homogenization to replace n by $n-1$ in Example 6.5.10. □

Corollary 8.4.4. *Let $n, d \in \mathbb{N}$. The following hold:*

(i) $\liminf_{d\to\infty} \dfrac{\mathcal{W}(n,2d)}{\dim \mathcal{H}_{n,2d}} \geq 1 - \dfrac{n-1}{2^{n-1}}$ *for all $n \in \mathbb{N}$.*

(ii) For any $d \geq 2$ and $\varepsilon > 0$ there exists a $n \in \mathbb{N}$ large enough such that

$$\mathcal{W}(n,2d) \geq (1-\varepsilon) \cdot \binom{n-1+2d}{n-1}.$$

[28] $\mathbb{R}$ replaced by an arbitrary closed field.

9. Tensor Decomposition

9.1. Definition

Definition 9.1.1. Let $d \in \mathbb{N}$, $\mathcal{V}$ be a vector space and let $\mathcal{V}^*$ be its dual. Then

$$(\mathcal{V}^*)^{\otimes d} := \underbrace{\mathcal{V}^* \otimes \cdots \otimes \mathcal{V}^*}_{d\text{-times}}$$

is the *d-tensor space* over $\mathcal{V}$. Let $w_1, \dots, w_d \in \mathcal{V}^*$ and $v_1, \dots, v_d \in \mathcal{V}$. Then we have the action on the *elementary tensor* $w_1 \otimes \cdots \otimes w_d$ given by

$$(w_1 \otimes \cdots \otimes w_d)(v_1, \dots, v_d) := w_1(v_1) \cdots w_d(v_d).$$

A tensor $t \in (\mathcal{V}^*)^{\otimes d}$ is called *symmetric* if

$$t(v_{\sigma(1)}, \dots, v_{\sigma(d)}) = t(v_1, \dots, v_d)$$

for any permutation σ of $\{1, \dots, d\}$.

Example 9.1.2. Let $n \in \mathbb{N}$. Then $\mathbb{R}^n \otimes \mathbb{R}^n \cong \mathbb{R}^{n \times n}$. For $w_1, w_2 \in \mathbb{R}^n$ the elementary tensor $w_1 \otimes w_2$ fulfills

$$(w_1 \otimes w_2)(v_1, v_2) = \langle w_1, v_1 \rangle \cdot \langle w_2, v_2 \rangle = (v_1^T \cdot w_1) \cdot (w_2^T \cdot v_2)$$

i.e., we have the identification $w_1 \otimes w_2 = w_1 \cdot w_2^T \in \mathbb{R}^{n \times n}$. ∘

9.2. Decomposition of Symmetric $2d$-Tensors

Question 9.2.1. *Given $w \in (\mathbb{R}^n)^{\otimes d}$ be a symmetric tensor. When can we write w as*

$$w = \sum_{i=1}^{k} w_i \otimes \cdots \otimes w_i? \tag{9}$$

How do we find the $w_1, \dots, w_d \in \mathbb{R}^n$ and the minimal $k \in \mathbb{N}_0$in (9)?

Theorem 9.2.2. *Let $n, d \in \mathbb{N}$. Then a symmetric tensor $w \in (\mathbb{R}^n)^{\otimes 2d}$ can be written as in (9) if and only if $w(x, \dots, x) \in \mathrm{War}(n, 2d) \subset \mathbb{R}[x_1, \dots, x_n]$ with $x = (x_1, \dots, x_n)$.*

Proof. Set $p(x) = w(x, \dots, x)$. The statement follows immediately from

$$(w_i \otimes \cdots \otimes w_i)(x, \dots, x) = \langle w_i, x \rangle^{2d} = (w_{i,1}x_1 + \cdots + w_{i,n}x_n)^{2d}.$$ □

Corollary 9.2.3. *The minimal k in (9) is the real Waring rank of w.*

10. Derivatives of Moments and Moment Functionals

The following section gives an overview of results presented in [dD19, dD23a].

10.1. Derivatives of Linear Functionals and Measures

Example 10.1.1. Let $a < b$ and $\chi_{[a,b]}$ be the characteristic function of the set $[a, b]$. Set μ as $\mathrm{d}\mu = \chi_{[a,b]}\,\mathrm{d}x$. Then the moments of μ are

$$s_k = \int_{\mathbb{R}} x^k \,\mathrm{d}\mu(x) = \int_a^b x^k \,\mathrm{d}x = \left[\frac{1}{k+1}x^{k+1}\right]_{x=a}^{b} = \frac{b^{k+1} - a^{k+1}}{k+1}.$$

But we of course have

$$\int_{\mathbb{R}} \partial_x x^k \,\mathrm{d}\mu(x) = \int_a^b k \cdot x^{k-1} \,\mathrm{d}x = [x^k]_{x=a}^b = b^k - a^k.$$

On the other hand let us assume we can do partial integration and we can do the following:

$$\int_{\mathbb{R}} \partial_x x^k \,\mathrm{d}\mu(x) = \int_{\mathbb{R}} (\partial_x x^k) \cdot \chi_{[a,b]} \,\mathrm{d}x = -\int_{\mathbb{R}} x^k \cdot \underbrace{\partial_x \chi_{[a,b]}\,\mathrm{d}x}_{\mathrm{d}(\delta_a - \delta_b)} = -(a^k - b^k) = b^k - a^k.$$

Here we take the derivative of a non-differentiable function, i.e.,

$$(\chi_{[a,b]})' = \delta_a - \delta_b$$

in the distributional sense [Gru09]. ◦

Definition 10.1.2. Let $n, d \in \mathbb{N}$ (or $d = \infty$) and $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ be a linear functional. For any $\alpha \in \mathbb{N}_0^n$ we define the *derivative* $\partial^\alpha L$ *of the linear functional* L by

$$\partial^\alpha L := (-1)^{|\alpha|} \cdot L \circ \partial^\alpha,$$

i.e., $(\partial^\alpha L)(p) = (-1)^{|\alpha|} \cdot L(\partial^\alpha p)$ for any $p \in \mathbb{R}[x_1, \dots, x_n]$. For any $\beta \in \mathbb{N}_0^n$ we define the *derivative* $\partial^\alpha s_\beta$ *of the moment* s_β by

$$\partial^\alpha s_\beta := (\partial^\alpha L)(x^\beta) = (-1)^{|\alpha|} \cdot L(\partial^\alpha x^\beta).$$

For a (moment) sequence $s = (s_\beta)_{\beta \in \mathbb{N}_0^n}$ we define the derivative $\partial^\alpha s := (\partial^\alpha s_\beta)_{\beta \in \mathbb{N}_0^n}$.

Lemma 10.1.3. *Let* $n \in \mathbb{N}$, $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ *be a moment functional, and* $\alpha \in \mathbb{N}_0^n \setminus \{0\}$. *The following are equivalent:*

(i) $\partial^\alpha L$ *is a moment functional.*

(ii) $\partial^\alpha L = 0$.

Proof. (ii) → (i) is clear. So we prove the inverse direction, i.e., let $\partial^\alpha L$ be a moment functional. Then $(\partial^\alpha L)(1) = (-1)^{|\alpha|} \cdot L(\partial^\alpha 1) = (-1)^{|\alpha|} \cdot L(0) = 0$. □

Definition 10.1.4. Let $n \in \mathbb{N}$, μ be a (signed) measure on $\mathbb{R}^n$ such that all moments are finite, and $\alpha \in \mathbb{N}_0^n$. If there exists a (signed) measure ν such that

$$\nu(f) = (-1)^{|\alpha|} \cdot \mu(\partial^\alpha f) \tag{10}$$

for all $f \in C_c^\infty(\mathbb{R}^n)$ then we say that the *derivative* $\partial^\alpha \mu := \nu$ *of the measure* μ exists.

Example 10.1.5. Derivatives of the Dirac measure δ_0 are not measures. For $k \in \mathbb{N}$ let $f_k(x) := \frac{1}{k} \cdot \sin(k^2 \cdot x)$. Then

$$(\delta_0')(f_k) = k \cdot \cos(0) = k \to \infty$$

as $k \to \infty$ which contradicts $\|f_k\|_\infty \to 0$ as $k \to \infty$. ○

Theorem 10.1.6. *Let $n \in \mathbb{N}$, $\alpha \in \mathbb{N}_0^n$, and $L : \mathbb{R}[x_1, \dots, x_n] \to \mathbb{R}$ be a moment functional with representing measure μ. If $\partial^\alpha \mu$ exists on $\mathbb{R}[x_1, \dots, x_n]$, then $\partial^\alpha \mu$ is a signed representing measure of $\partial^\alpha L$, i.e.,*

$$s_\beta = \mu(x^\beta) \qquad \Rightarrow \qquad (\partial^\alpha \mu)(x^\beta) = \partial^\alpha s_\beta$$

for all $\beta \in \mathbb{N}_0^n$.

Proof. Since $\partial^\alpha \mu$ exists on $\mathbb{R}[x_1, \dots, x_n]$ we have

$$(\partial^\alpha \mu)(x^\beta) \overset{\text{Def. 10.1.4}}{=} \nu(x^\beta) \overset{(10)}{=} \mu(\partial^\alpha x^\beta) \overset{\text{Def. 10.1.2}}{=} \partial^\alpha s_\beta. \qquad \square$$

Besides the Dirac measures also measures of the form $f \, \mathrm{d}\lambda^n$ are very important, where λ^n is the n-dimensional Lebesgue measure and f is a measurable function.

Definition 10.1.7 ([Gru09, Eq. (3.2)])**.** Let $f \in L^1_{\mathrm{loc}}(\mathcal{X})$ and λ^n the n-dimensional Lebesgue measure on $\mathcal{X}$. We define the distribution Λ_f by

$$\Lambda_f(g) \quad := \quad \int_{\mathcal{X}} g(x) f(x) \, \mathrm{d}\lambda^n(x), \qquad \text{for all } g \in \mathcal{D}(\mathcal{X}).$$

Theorem 10.1.8 ([Gru09, Eqs. (3.15) and (3.21)])**.** *Let $\alpha \in \mathbb{N}_0^n$. Then*

$$\partial^\alpha \Lambda_f = \Lambda_{\partial^\alpha f}, \qquad \textit{for all } f \in L^1_{\mathrm{loc}}(\mathcal{X}). \tag{11}$$

10.2. Polytope Reconstruction

The problem of reconstructing a (convex and full-dimensional) polytope $P \subset \mathbb{R}^n$, i.e., finding all vertices, is an extensively studied question and several algorithms have been proposed, see e.g. [Bal61, MN68, MR80, LR82, MVKW95, GMV99, BGL07, GLPR12, GNPR14, GPSS18, KSS18], and references therein.

Based on derivatives of moments we will present a simple proof of one version of these algorithms which calculates the vertices from finitely many moments

$$s_\alpha = \int x^\alpha \cdot \chi_P \, \mathrm{d}\lambda^n(x).$$

We use the Brion–Lawrence–Khovanskii–Pukhlikov–Barvinok (BBaKLP) formulas [Bri88, Law91, Bar91, PK92, Bar92] and the generalized eigenvalue problem.

Let us state the BBaKLP formulas. This presentation is taken from [GLPR12]. Let P be a polytope in $\mathbb{R}^n$ with vertices $v_1, \dots, v_k$ ($k \geq n+1$), then

$$0 \quad = \quad \sum_{i=1}^{k} \langle v_i, r\rangle^j \tilde{D}_{v_i}(r) \qquad \text{for all } j = 0, \dots, n-1, \tag{12}$$

see [GLPR12, Eq. (3)], and for $j = n, n+1, \dots$ we have

$$\int_P \langle x, r\rangle^j \, \mathrm{d}^n x \quad =: \quad s_j(r) \quad = \quad \frac{j!(-1)^n}{(j+n)!} \sum_{i=1}^{k} \langle v_i, r\rangle^{j+n} \tilde{D}_{v_i}(r), \tag{13}$$

see [GLPR12, Eq. (4)], where $\tilde{D}_{v_i}(r)$ is a rational function on $r \in \mathbb{R}^n$, i.e., r can be chosen in general position such that $\tilde{D}_{v_i}(\,\cdot\,)$ has no zero or pole at r. The $s_j(r)$ is the *j-th directional moment* with direction r.

Definition 10.2.1. Let $k, n \in \mathbb{N}$, P be a polytope with vertices $v_1, \dots, v_k \in \mathbb{R}^n$, $r \in \mathbb{R}^n \setminus \{0\}$ a vector (of length 1), $a \in \mathbb{R}$, and $H_{r,a} := \{x \in \mathbb{R}^n \mid \langle r, x\rangle = a\}$ be an affine hyperplane with normal vector r. We define the *area function* $\Theta_{P,r}$ to be the $(n-1)$-dimensional volume of $P \cap H_{r,x}$

$$\Theta_{P,r} : \mathbb{R} \to \mathbb{R},\ x \mapsto \Theta_{P,r}(x) := \mathrm{vol}_{n-1}(P \cap H_{r,x}) = \int_{H_{r,x}} \chi_P(y) \, \mathrm{d}\lambda^{n-1}(y)$$

where λ^{n-1} is the $(n-1)$-dimensional Lebesgue measure on $H_{r,x}$.

Of course, the area function is integration by parts

$$s_j(r) \quad = \quad \int_{\mathbb{R}^n} \langle x, r\rangle^j \cdot \chi_P \, \mathrm{d}\lambda^n(x) \quad = \quad \int_{\mathbb{R}} y^j \cdot \Theta_{P,r}(y) \, \mathrm{d}\lambda(y).$$

The area function $\Theta_{P,r}$ is a continuous piecewise polynomial function of degree n if r is not a normal vector of any facet of P.

Lemma 10.2.2. *Let $r \in \mathbb{R}^n$ be a vector of unit length such that $\tilde{D}_{v_i}(r)$ is non-zero and well-defined, i.e., its numerator and denominator is non-zero. Then*

$$\partial^n \Lambda_{\Theta_{P,r}} \quad = \quad \sum_{i=1}^{k} \tilde{D}_{v_i}(r) \cdot \delta_{\langle r, v_i\rangle}. \tag{14}$$

Proof. Set $y := \langle x, r\rangle$. From (12) for $j = 0, \dots, n-1$ we have

$$\int y^j \cdot \partial^n \Theta_{P,r}(y) \, \mathrm{d}y \overset{(*)}{=} (-1)^n \int \partial^n y^j \cdot \Theta_{P,r}(y) \, \mathrm{d}y = 0 = \sum_{i=1}^{k} \langle v_i, r\rangle^j \tilde{D}_{v_i}(r)$$

and from (13) with $j' \geq 0$ we have

$$\int y^{n+j'} \cdot \partial^n \Theta_{P,r}(y)\ \mathrm{d}y \overset{(+)}{=} (-1)^n \int \partial^n y^{n+j'} \cdot \Theta_{P,r}(y)\ \mathrm{d}y$$
$$= \frac{(-1)^n (n+j')!}{j'!} \int y^{j'} \cdot \Theta_{P,r}(y)\ \mathrm{d}y = \sum_{i=1}^k \langle v_i, r\rangle^{j'+n} \tilde{D}_{v_i}(r).$$

Here $(*)$ and $(+)$ hold since $\operatorname{supp} \Theta_{P,r}$ is compact. Thus the claim follows since the set of polynomial functions on a compact set K is dense in $C^\infty(K)$. $\square$

In the previous proof the BBaKLP formulas were used for all monomials y^j ($j \in \mathbb{N}_0$) and the Weiserstraß Theorem gives the assertion. But the proof of the lemma can be weakened to the Müntz–Szász Theorem [Mün14, Szá16], i.e., only monomials $\{y^{d_i}\}_{i\in\mathbb{N}}$ with $\sum_{i\in\in\mathbb{N}} \frac{1}{d_i} = \infty$ (and $d_1 = 0$) are necessary. Additionally, the BBaKLP formulas hold only for polynomials but the previous lemma applies to all C^n-functions. So we have the following.

Theorem 10.2.3. *Let $\mathcal{A}$ be a (finite-dimensional) vector space of measurable functions on $\mathbb{R}$ with basis $\mathsf{A} = \{a_1, a_2, \dots\}$ such that $\partial\mathcal{A} \subseteq \mathcal{A}$, i.e., $\partial^d \mathcal{A} \subseteq \mathcal{A}$ for all $d \in \mathbb{N}$. Let $P \subset \mathbb{R}^n$ be a polytope with vertices $v_1, \dots, v_k$, $k \geq n+1$, $r \in \mathbb{R}^n$ be such that it is neither a pole nor a zero of any $\tilde{D}_{v_i}(\,\cdot\,)$, and consider the directional moments*

$$s_j \quad = \quad s_j(r) \quad := \quad \int_P a_j(\langle x, r\rangle)\ \mathrm{d}\lambda^n(x).$$

Then $\partial^n s$ has an at most k-atomic signed representing measure

$$\partial^n \Lambda_{\Theta_{P,r}} \quad = \quad \sum_{i=1}^k \tilde{D}_{v_i}(r) \cdot \delta_{\langle v_i, r\rangle}$$

supported only at the projections $\langle v_i, r\rangle$ of the vertices v_i.

Proof. Since s has the representing measure $\Lambda_{\Theta_{P,r}}$, the $\partial^n s$ has the at most k-atomic representing (signed) measure $\partial^n \Lambda_{\Theta_{P,r}} = \sum_{i=1}^k \tilde{D}_{v_i}(r) \cdot \delta_{\langle v_i, r\rangle}$ by Theorem 10.1.6 and Lemma 10.2.2. $\square$

Corollary 10.2.4. *Let $P \subset \mathbb{R}^n$ be a polytope with vertices $v_1, \dots, v_k$, $k \geq n+1$ and let $r \in \mathbb{R}^n$ be such that it is neither a pole nor a zero of any $\tilde{D}_{v_i}(\,\cdot\,)$, and for $j = 0, \dots, 2k-n+1$ let $s_j = s_j(r)$ be the directional moments*

$$s_j \quad = \quad \int_P \langle x, r\rangle^j\ \mathrm{d}\lambda^n(x).$$

Then the projections $\xi_i := \langle v_i, r\rangle$ are the eigenvalues of the generalized eigenvalue problem

$$\mathcal{H}_k(M_1 \partial^n s) y_i \quad = \quad \xi_i \cdot \mathcal{H}_k(\partial^n s) y_i. \tag{15}$$

Proof. As in Theorem 10.2.3 $s = (s_i)_{i=0}^{2k+1}$ has the representing measure $\Lambda_{\Theta_{P,r}}$ and $\partial^n s$ has the at most k-atomic representing (signed) measure $\partial^n \Lambda_{\Theta_{P,r}} = \sum_{i=1}^k \tilde{D}_{v_i}(r) \cdot \delta_{\langle v_i, r\rangle}$ by Theorem 10.1.6 and Lemma 10.2.2. By Theorem 7.2.3 the positions $\xi_i = \langle v_i, r\rangle$ are the eigenvalues of the generalized eigenvalue problem (15). $\square$

Remark 10.2.5. In [GLPR12, Eq. (5)] a "scaled vector of moments" is defined in a similar way as $\partial^n s$. However, the strength of Theorem 10.1.6, in particular in combination with Theorem 10.1.8, has not been used. $\circ$

Remark 10.2.6. With $n+1$ different directions r the vertices can be reconstructed using the previous theorem and $(n+1)(2k-n)+1$ moments are required. If k is unknown, the previous theorem also determines k if sufficiently many directional moments are given. $\circ$

Now we extend Definition 10.2.1 to functions f:

$$\Theta_{f,r}(x) := \int_{H_{r,x}} f(y)\ \mathrm{d}\lambda^{n-1}(y), \tag{16}$$

i.e., integration by part over $H_{r,x}$.

By linearity of integration and differentiation Corollary 10.2.4 also detects the vertices $v_{i,j}$, $j = 1, \dots, d_i$, of full-dimensional polytopes $P_i \subset \mathbb{R}^n$, $j = 1, \dots, p$, from the moments

$$s_k(r) := \int_{\mathbb{R}^n} \langle x, r\rangle^k \cdot \chi(x)\ \mathrm{d}\lambda^n(x) \tag{17}$$

of the simple function

$$\chi := \sum_{i=1}^p c_i \cdot \chi_{P_i} \qquad (c_i \in \mathbb{R},\ c_i \neq 0) \tag{18}$$

if the P_i or c_i are in general position. We say that a set $\{P_i\}_{i=1}^p$ of polytopes is *in general position* iff $v_{i,j} \neq v_{i',j'}$ for all $(i,j) \neq (i',j')$. Furthermore, we say that $c_1, \dots, c_p$ are *in general position* iff

$$\mu = \sum_{i=1}^p \sum_{j=1}^{d_i} c_i \cdot \tilde{D}_{v_{i,j}}(r) \cdot \delta_{\langle v_{i,j}, r\rangle} \tag{19}$$

has non-zero mass $\mu(\langle v_{i,j}, r\rangle) \neq 0$ for $r \in \mathbb{R}^n$ in general position, i.e., coefficients in (19) do not cancel out for vertices $v_{i,j}$ with the same projection $\langle v_{i,j}, r\rangle$.

Theorem 10.2.7. *Let $P_i \subset \mathbb{R}^n$, $i = 1, \dots, p$, be full-dimensional polytopes with vertices $v_{i,j}$, $j = 1, \dots, d_i$. Let the vertices $v_{i,j}$ or $c_1, \dots, c_p$ be in general position. Let $d := d_1 + \dots + d_p$. Then for a direction $r \in \mathbb{R}^n$ in general position the projections $\xi_{i,j} := \langle r, v_{i,j}\rangle$ are the eigenvalues of the generalized eigenvalue problem*

$$\mathcal{H}_d(M_1 \partial^n s) y_{i,j} = \xi_{i,j} \mathcal{H}_d(\partial^n s) y_{i,j} \tag{20}$$

where $s_0, \dots, s_{2d-n+1}$ are the directional moments (17) of (18).

Proof. By linearity of ∂^n and Lemma 10.2.2 we have that

$$\partial^n \Lambda_{\Theta_{\xi,r}} = \sum_{i=1}^{p} c_i \cdot \partial^n \Lambda_{\Theta_{P_i,r}} = \sum_{i=1}^{p} \sum_{j=1}^{d_i} c_i \cdot \tilde{D}_{v_{i,j}}(r) \cdot \delta_{\langle v_{i,j},r\rangle}$$

is a (signed) representing measure of $\partial^n s$ (Theorem 10.1.6). Then $(\partial^n \Lambda_{\Theta_{\xi,r}})(\langle r, v_{i,j}\rangle) \neq 0$ for all i,j since the $v_{i,j}$ or c_i are in general position. Hence the projections $\langle r, v_{i,j}\rangle$ are the eigenvalues of (20) by Theorem 7.2.3. $\square$

11. Gaussian Distributions and Mixtures

11.1. Reconstruction of One Component

For a Gaussian distribution $g(x) = c \cdot \exp(-\frac{a}{2}(x-b)^2)$ on $\mathbb{R}$ we have

$$g'(x) \;=\; -a(x-b)\cdot g(x) \;=\; -ax\cdot g(x) + ab\cdot g(x). \tag{21}$$

So integration over $x^i \cdot g'(x)$ gives

$$-i\cdot s_{i-1} \;=\; (\partial s)_i \;=\; -a\cdot (M_1 s)_i + ab\cdot s_i \;=\; -as_{i+1} + ab\cdot s_i, \quad \text{for all } i\in\mathbb{N}_0, \tag{22}$$

see also [AFS16, Eq. (5)]. This implies the following result.

Lemma 11.1.1 ([AFS16, Prop. 1])**.** *Let $k \in \mathbb{N}$, $k \geq 2$, be a natural number and $s = (s_0, s_1, \dots, s_k)$ be a real sequence with $s_0 \neq 0$. The following are equivalent:*

i) s is the moment sequence of the Gaussian distribution $c\cdot\exp(-\frac{a}{2}(x-b)^2)$ with $a,b,c\in\mathbb{R}$, $a>0$, $c\neq 0$, i.e., $s_i = \int x^i \cdot c \cdot \exp(-\frac{a}{2}(x-b)^2)\,\mathrm{d}x$.

ii) There are $a,b\in\mathbb{R}$ with $a>0$ such that the matrix

$$(\partial s, s, M_1 s)_{k-1} = \begin{pmatrix} 0 & s_0 & s_1 \\ -s_0 & s_1 & s_2 \\ -2\cdot s_1 & s_2 & s_3 \\ \vdots & \vdots & \vdots \\ -(k-1)\cdot s_{k-2} & s_{k-1} & s_k \end{pmatrix}$$

has rank two with kernel $(1, -ab, a)^T \cdot \mathbb{R}$.

In this case, one has $a = \frac{s_0^2}{s_0 s_2 - s_1^2}$, $b = \frac{s_1}{s_0}$ and $c = s_0 \cdot \sqrt{\frac{a}{\pi}}$.

Proof. While (i) $\Rightarrow$ (ii) is clear, we show (ii) $\Rightarrow$ (i) by induction on i. Since $0 \neq s_0 = c \cdot \int e^{-a(x-b)^2}\,\mathrm{d}x$ for $c = s_0 \cdot \sqrt{\frac{a}{\pi}}$ and $s_{-1} := 0$, we have by (ii), (21), (22) and the induction hypothesis that

$$\begin{aligned}
a \cdot s_{i+1} &= i \cdot s_{i-1} + ab \cdot s_i \\
&= \int \partial x^i \cdot c \cdot \exp(-a(x-b)^2)\,\mathrm{d}x + \int ab \cdot x^i \cdot c \cdot \exp(-a(x-b)^2)\,\mathrm{d}x \\
&= \int [-x^i(-ax+ab) + ab \cdot x^i] \cdot c \cdot \exp(-a(x-b)^2)\,\mathrm{d}x \\
&= a \cdot \int x^{i+1} \cdot c \cdot \exp(-a(x-b)^2)\,\mathrm{d}x \qquad \text{for all } i = 0,\dots,k-1,
\end{aligned}$$

i.e., s_{i+1} is the $(i+1)$-th moment of $c \cdot \exp(-a(x-b)^2)$. $\square$

On $\mathbb{R}^n$ we have the following.

Theorem 11.1.2. *Let $n \in \mathbb{N}$, $A = (a_1,\dots,a_n) = (a_{i,j})_{i,j=1}^n \in \mathbb{R}^{n\times n}$ be a symmetric and positive definite matrix, $b \in \mathbb{R}^n$, $c \in \mathbb{R}$, $c \neq 0$, and $k \in \mathbb{N}$ with $k \geq 2$. Set*

$$g(x) := c \cdot e^{-\frac{1}{2}(x-b)^T A (x-b)}.$$

For a multi-indexed real sequence $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\leq k}$ the following are equivalent:

i) s is the moment sequence of Λ_g, i.e., $s_\alpha = \int x^\alpha \cdot g(x)\,\mathrm{d}\lambda^n(x)$ for all $\alpha \in \mathbb{N}_0^n$ with $|\alpha| \leq k$.

ii) For $i = 1,\dots,n$ the matrix $(\partial_i s, s, M_{e_1}s, \dots M_{e_n}s)_{k-1}$ has the 1-dimensional kernel

$$(1, -\langle b, a_i\rangle, a_{i,1}, \dots, a_{i,n}) \cdot \mathbb{R}. \tag{23}$$

Proof. For $i = 1,\dots,n$ we have

$$0 = \partial_i g(x) - \langle b, a_i\rangle \cdot g(x) + a_{i,1}x_1 \cdot g(x) + \dots + a_{i,n}x_n \cdot g(x). \tag{$*$}$$

(i) $\Rightarrow$ (ii): From $(*)$ we find that (23) is contained in the kernel of the matrix $(\partial_i s, s, M_{e_1}s, \dots, M_{e_n}s)_{k-1}$. It suffices to show that the kernel of the matrix $(\partial_i s, s, M_{e_1}s, \dots, M_{e_n}s)_1$ is at most one-dimensional. Consider

$$H := \begin{pmatrix} s_0 & s_{e_1} & \dots & s_{e_n} \\ s_{e_1} & s_{2e_1} & \dots & s_{e_1+e_n} \\ \vdots & \vdots & & \vdots \\ s_{e_n} & s_{e_1+e_n} & \dots & s_{2e_n} \end{pmatrix},$$

the Hankel matrix of $L_s|_{\mathbb{R}[x_1,\dots,x_n]_{\leq 2}}$. Let $d = (d_0,\dots,d_n) \in \ker H$. Then $0 = L_s(\langle d, (1,x_1,\dots,x_n)\rangle^2) = \int (d_0 + d_1x_1 + \dots + d_nx_n)^2\,\mathrm{d}\Lambda_g(x)$ implies $d = 0$, i.e., H has full rank $n+1$. Therefore

$(\partial_i s, s, M_{e_1} s, \dots, M_{e_n} s)_1$ has rank at least $n+1$ since it has H as submatrix. Its kernel can thus be at most one-dimensional.

(ii) $\Rightarrow$ (i): Let $O \in \mathbb{R}^{n\times n}$ be an orthogonal matrix such that $O\cdot A\cdot O^T = \operatorname{diag}(\lambda_1,\dots,\lambda_n)$, $\lambda_i > 0$. The coordinate change on $\mathbb{R}^n$ given by $y = Ox$ induces a linear transformation on the space of moment sequences. Let $t = (t_\alpha)_{|\alpha|\leq k}$ be the moment sequence obtained from s via this transformation. A straight-forward calculation shows that

$$\begin{aligned}\ker(\partial_i t, t, M_{e_1} t, \dots, M_{e_n} t)_1 &= \ker(\partial_i t, t, M_{e_1} t, \dots, M_{e_n} t)_{k-1}\\ &= (1, -\lambda_i \tilde{b}_i, 0, \dots, 0, \lambda_i, 0, \dots, 0)^T \cdot \mathbb{R},\end{aligned}$$

where $\tilde{b} = Ob$. This means that we are in the 1-dimensional setting

$$\ker(\partial_i (t_{j\cdot e_i})_{j=1}^k, (t_{j\cdot e_i})_{j=1}^k, M_{e_i} (t_{j\cdot e_i})_{j=1}^k) = (1, -\lambda_i \tilde{b}_i, \lambda_i)^T \cdot \mathbb{R}$$

where the 1-dimensional assertion holds by Lemma 11.1.1. Hence, $t = (t_\beta)$ is represented by $t_0 \cdot \frac{\sqrt{\lambda_1\cdots\lambda_n}}{(\pi)^{n/2}} \prod_{i=1}^n e^{-\frac{\lambda_i}{2}(y_i - \tilde{b}_i)^2}$. The inverse transformation $x = O^T y$ together with $\lambda_1\cdots\lambda_n = \det(A)$ gives the n-dimensional assertion. □

Hence, the previous theorem provides an easy way to determine $A \in \mathbb{R}^{n\times n}$ and $b \in \mathbb{R}^n$ from the moments s_α.

Algorithm 11.1.3.

Input: *$k \in \mathbb{N}$, $k \geq 2$; $s = (s_\alpha)_{\alpha\in\mathbb{N}_0^n : |\alpha|\leq k}$.*

Step 1: For $i = 1,\dots,n$:

a) Calculate β_i and $a_i = (a_{i,1},\dots,a_{i,n})$ from

$$\ker(\partial_i s, s, M_{e_1} s, \dots M_{e_n} s)_1 = (1, -\beta_i, a_{i,1}, \dots, a_{i,n}) \cdot \mathbb{R}$$

- *If the kernel is not one-dimensional, then s is not represented by one Gaussian distribution.*

b) Check: $(1, -\beta_i, a_{i,1}, \dots, a_{i,n}) \in \ker(\partial_i s, s, M_{e_1} s, \dots M_{e_n} s)_{k-1}$?

- *If FALSE: s is not represented by one Gaussian distribution.*

Step 2: Check: $A = (a_{i,j})_{i,j=1}^n$ is symmetric and positive definite?

- *If FALSE: s is not represented by one Gaussian distribution.*

Step 3: Calculate $b = A^{-1}\cdot(\beta_1,\dots,\beta_n)^T$ and $c = \frac{\sqrt{\det(A)}}{\pi^{n/2}} \cdot s_0$.

Out: *"s is represented by a Gaussian distribution": TRUE or FALSE. If TRUE: A, b, c.*

11.2. Lower Bounds on Components in a Gaussian Mixture from finitely many Moments

Theorem 11.2.1. *Let $n, d \in \mathbb{N}$. Let $p \in \mathbb{R}[x_1, \dots, x_n]_{\leq 2d}$ be a non-negative polynomials with finitely many zeros $z_1, \dots, z_k$. Then there exists a moment sequence $s \in \operatorname{int} \mathcal{S}_\mathsf{A}$ and an open neighborhood of s contained in the moment cone $\mathcal{S}_\mathsf{A}$ such that s has a Gaussian mixture representation which needs at least*

$$\dim \operatorname{lin} \{ s_\mathsf{A}(z_i) \mid i = 1, \dots, k \}$$

many Gaussians but not less. All variance matrices can be chosen to be equal and dropping this restriction does not reduce the number of needed Gaussians.

Proof. See [dD23a]. □

Corollary 11.2.2. *Let $d \in \mathbb{N}$ and $\varepsilon > 0$. Then there is an $n \in \mathbb{N}$ such that there is a moment functional $L : \mathbb{R}[x_1, \dots, x_n]_{\leq 2d} \to \mathbb{R}$ which can be written as a sum of*

$$(1 - \varepsilon) \cdot \binom{2d + n}{n}$$

Gaussian distributions but not less.

12. Moments and Partial Differential Equation

12.1. The Heat Equation acting on Moment Sequences: Gaussian Mixtures

The results presented here are published in [CdD22].

Definition 12.1.1. Let $n \in \mathbb{N}$. The set of *Schwartz functions* $\mathcal{S}(\mathbb{R}^n)$ is defined as

$$\mathcal{S}(\mathbb{R}^n) := \{ f \in C^\infty(\mathbb{R}^n, \mathbb{R}) \mid \| x^\alpha \cdot \partial^\beta f(x) \|_\infty < \infty \text{ for all } \alpha, \beta \in \mathbb{N}_0^n \}.$$

Theorem 12.1.2. *Let $n \in \mathbb{N}$ and $u_0 \in \mathcal{S}(\mathbb{R}^n)$. Then the* heat equation

$$\begin{aligned} \partial_t u(x,t) &= \Delta u(x,t) \\ u(x,0) &= u_0 \end{aligned}$$

has the unique solution $u(x,t) = (\Theta_t * u_0)(x) \in C^\infty([0,\infty), \mathcal{S}(\mathbb{R}^n))$ where Θ_t is the heat kernel

$$\Theta_t(x) := \frac{1}{(4\pi t)^{n/2}} \cdot e^{-\frac{x^2}{4t}}$$

and $*$ denotes the convolution. If additionally $u_0 \geq 0$, then $u(\,\cdot\,, t) \geq 0$ for all $t \geq 0$.

For more on the heat equation see, e.g., [Eva10, Ch. 2.3]. For $u_0 \in \mathcal{S}(\mathbb{R}^n)$ all moments of the measure $u(x,t)\ \mathrm{d}x$ exist and are time-dependent.

Definition 12.1.3. Let $n \in \mathbb{N}$, $u_0 \in \mathcal{S}(\mathbb{R}^n)$ with $u_0 \geq 0$, and u be the solution of the heat equation. We define the *time-dependent moment from the heat equation by*

$$s_\alpha(t) := \int_{\mathbb{R}^n} x^\alpha \cdot u(x,t)\, \mathrm{d}x \quad \text{with} \quad \alpha \in \mathbb{N}_0^n.$$

Therefore, $s(t) := (s_\alpha(t))_{\alpha\in\mathbb{N}_0^n}$ is a moment sequence with representing measure μ_t with $\mathrm{d}\mu_t(x) = u(x,t)\, \mathrm{d}x$ for all $t \geq 0$. The 1-parameter family $s(t)$ of moment sequences describes a curve in the moment cone.

Lemma 12.1.4. *Let $n \in \mathbb{N}$, $u_0 \in \mathcal{S}(\mathbb{R}^n)$, and u be a solution of the heat equation. Then the moments $s_\alpha(t)$ of $u(x,t)\, \mathrm{d}x$ fulfill the following.*

(i) *$s_\alpha \in \mathbb{R}[t]$ with*

$$\deg s_\alpha \leq \left\lfloor \frac{\alpha_1}{2} \right\rfloor + \dots + \left\lfloor \frac{\alpha_n}{2} \right\rfloor$$

for all $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{N}_0^n$. The coefficients of s_α only depend on the moments $s_\beta(0)$ of u_0 with $\beta \leq \alpha$, i.e., $\beta_j \leq \alpha_j$ for all $j = 1,\dots,n$.

(ii) *For $n = 1$ and all $k \in \mathbb{N}_0$, we have*

$$s_{2k}(t) = \sum_{j=0}^{k} \frac{(2k)!}{(2k-2j)! \cdot j!} \cdot s_{2k-2j}(0) \cdot t^j$$

and

$$s_{2k+1}(t) = \sum_{j=0}^{k} \frac{(2k+1)!}{(2k+1-2j)! \cdot j!} \cdot s_{2k+1-2j}(0) \cdot t^j$$

Proof. (i): Let $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{N}_0^n$. Since u solves the heat equation, we have

$$\begin{aligned}\partial_t s_\alpha(t) &= \partial_t \int_{\mathbb{R}^n} x^\alpha \cdot u(x,t)\, \mathrm{d}x = \int_{\mathbb{R}^n} x^\alpha \cdot \partial_t u(x,t)\, \mathrm{d}x\\ &= \int_{\mathbb{R}^n} x^\alpha \cdot \Delta u(x,t)\, \mathrm{d}x,\end{aligned}$$

and since $u(\,\cdot\,,t) \in \mathcal{S}(\mathbb{R}^n)$ for all $t \geq 0$, we apply partial integration

$$\begin{aligned}&= \sum_{j=1}^{n} \int_{\mathbb{R}^n} x^\alpha \cdot \partial_j^2 u(x,t)\, \mathrm{d}x = -\sum_{j=1}^{n} \int_{\mathbb{R}^n} (\partial_j x^\alpha) \cdot \partial_j u(x,t)\, \mathrm{d}x\\ &= \sum_{j=1}^{n} \int_{\mathbb{R}^n} (\partial_j^2 x^\alpha) \cdot u(x,t)\, \mathrm{d}x = \int_{\mathbb{R}^n} (\Delta x^\alpha) \cdot u(x,t)\, \mathrm{d}x\end{aligned}$$

where $\partial_j := \frac{\partial}{\partial x_j}$. With e_j denoting the n-tuple with j-th coordinate equal to 1 and zeros elsewhere, this implies

$$\Delta x^\alpha = (\partial_1^2 + \dots + \partial_n^2)x^\alpha = \sum_{j=1}^n \alpha_j \cdot (\alpha_j - 1) \cdot x^{\alpha - 2e_j}.$$

This gives

$$\partial_t s_\alpha(t) = \sum_{j=1}^n \alpha_j \cdot (\alpha_j - 1) \cdot s_{\alpha - 2e_j}(t) \tag{24}$$

with initial values $s_\alpha(0)$ and with $s_{\alpha-2e_j}(t) = 0$ for all α with $\alpha_j \leq 1$ for some $j = 1, \dots, n$. Now observe that (24) is a recursive system of ODEs. We proceed by induction. Let $\alpha = (\alpha_1, \dots, \alpha_n)$ with $\alpha_j \in \{0, 1\}$. Then the degree bound holds since $\partial_t s_\alpha(t) = 0$, i.e., $s_\alpha(t) = s_\alpha(0)$. Hence, integrating (24) gives

$$\begin{aligned}
\deg s_\alpha(t) &= 1 + \max_{j=1,\dots,n} \deg s_{\alpha-2e_j}(t) \\
&\leq 1 + \max_{j=1,\dots,n} \left(\left\lfloor \frac{\alpha_1}{2} \right\rfloor + \dots + \left\lfloor \frac{\alpha_j - 2}{2} \right\rfloor + \dots + \left\lfloor \frac{\alpha_n}{2} \right\rfloor \right) \\
&= \left\lfloor \frac{\alpha_1}{2} \right\rfloor + \dots + \left\lfloor \frac{\alpha_j}{2} \right\rfloor + \dots + \left\lfloor \frac{\alpha_n}{2} \right\rfloor.
\end{aligned}$$

(ii): From (24) we get $\partial_t s_0(t) = 0$, i.e., $s_0(t) = s_0(0)$, and

$$\partial_t s_{2j}(t) = 2j \cdot (2j-1) \cdot s_{2j-2}(t) \tag{25}$$

as well as $\partial_t s_1(t) = 0$, i.e. $s_1(t) = s_1(0)$, and

$$\partial_t s_{2j+1}(t) = (2j+1) \cdot 2j \cdot s_{2j-1}(t) \tag{26}$$

(25) and (26) can easily be solved by recursion, thus establishing (ii). □

Definition 12.1.5. Let $n \in \mathbb{N}$ and $d \in \mathbb{N} \cup \{\infty\}$. For a sequence $s = (s_\alpha(0))_{\alpha \in \mathbb{N}_0^n : |\alpha| \leq d}$, $|\alpha| := \alpha_1 + \dots + \alpha_n$, we define

$$\mathfrak{p}_s := (\mathfrak{p}_{s,\alpha})_{\alpha \in \mathbb{N}_0^n : |\alpha| \leq d} \quad \subset \mathbb{R}[t]$$

where $\mathfrak{p}_{s,\alpha}$ are the polynomials $s_\alpha(t)$ as in Lemma 12.1.4.

Corollary 12.1.6. *Let s, s' be (real) sequences and $a, b, t_1, t_2 \in \mathbb{R}$. $\mathfrak{p}$ has the following properties:*

(i) $\mathfrak{p}_s(0) = s$,

(ii) $\mathfrak{p}_{\mathfrak{p}_s(t_1)}(t_2) = \mathfrak{p}_s(t_1 + t_2)$, *and*

(iii) $\mathfrak{p}_{a \cdot s + b \cdot s'} = a \cdot \mathfrak{p}_s + b \cdot \mathfrak{p}_{s'}$.

Proof. (i) and (iii) are clear. (ii) follows from the semi-group property of the heat kernel: $\Theta_{t_1} * (\Theta_{t_2} * u_0) = \Theta_{t_1+t_2} * u_0$. □

Example 12.1.7. For $n = 1$, $k \in \mathbb{N}_0$, and $\mathfrak{p}_{s,k}$ the polynomials in Lemma 12.1.4(ii) we have

$$\begin{aligned}
\mathfrak{p}_{s,0}(t) &= s_0(t) = s_0(0)\\
\mathfrak{p}_{s,1}(t) &= s_1(t) = s_1(0)\\
\mathfrak{p}_{s,2}(t) &= s_2(t) = s_2(0) + 2s_0(0)\cdot t\\
\mathfrak{p}_{s,3}(t) &= s_3(t) = s_3(0) + 6s_1(0)\cdot t\\
\mathfrak{p}_{s,4}(t) &= s_4(t) = s_4(0) + 12s_2(0)\cdot t + 12s_0(0)\cdot t^2\\
\mathfrak{p}_{s,5}(t) &= s_5(t) = s_5(0) + 20s_3(0)\cdot t + 60s_1(0)\cdot t^2\\
&\;\;\vdots
\end{aligned} \tag{27}$$

○

Theorem 12.1.8. *Let $k \in \mathbb{N}$, $N \in \mathbb{N}\cup\{\infty\}$, $p_1,\dots,p_k \in \mathbb{R}^n$, $c_1,\dots,c_k \in \mathbb{R}$, and $t_1,\dots,t_k \in [0,\infty)$. Set $\tau := \min_i t_i$ and $s := (s_\alpha)_{\alpha\in\mathbb{N}_0^n:|\alpha|\le N}$ with*

$$s_\alpha := \int_{\mathbb{R}^n} x^\alpha\; \mathrm{d}\mu_0(x) \qquad \textit{and} \qquad \mu_0(x) := \sum_{i=0}^k c_i\cdot\Theta_{t_i}(x-p_i).$$

Then for all $t \in [-\tau,\infty)$ we have that the sequence $\mathfrak{p}_s(t)$ is represented by

$$\mu_t(x) := \sum_{i=0}^k c_i\cdot\Theta_{t_i+t}(x-p_i).$$

Proof. By Lemma 12.1.4 and the linearity of the integral (moments), in the measure $\mu_t(x)$, it is sufficient to show the statement for $k = 1$. Now, for $t \in (-t_1, 0)$, the statement follows from the semi-group property of the heat kernel, i.e., $\Theta_t * \Theta_{t_1-t} = \Theta_{t_1}$; for $t \in (0,\infty)$, it follows from $\Theta_t * \Theta_{t_1} = \Theta_{t_1+t}$. It remains to treat the special case $t = -t_1$. This case follows from

$$\int_{\mathbb{R}^n} x^\alpha\cdot\Theta_t(x-x_0)\;\mathrm{d}x \quad\xrightarrow{t\to 0}\quad \int_{\mathbb{R}^n} x^\alpha\;\mathrm{d}\delta_{x_0}(x) = x_0^\alpha$$

for all $\alpha \in \mathbb{N}_0^n$. □

Definition 12.1.9. Let $n \in \mathbb{N}$ and $d \in \mathbb{N}\cup\{\infty\}$. For $s \in \mathcal{S}_d$ we define

$$\mathfrak{I}_s := \{t \in \mathbb{R} \,|\, \mathfrak{p}_s(t) \in \mathcal{S}_d\}.$$

Theorem 12.1.10. *Let $n \in \mathbb{N}$, $d \ge 2$ or $d = \infty$ and $s \in \mathcal{S}_d \setminus \{0\}$. Then*

$$\mathfrak{I}_s = [-\mathfrak{d}_s, \infty) \quad \textit{or} \quad (-\mathfrak{d}_s,\infty)$$

with

$$\mathfrak{d}_s \in \left[0, \frac{s_{2e_1} + \cdots + s_{2e_n}}{2n \cdot s_0}\right].$$

For $d = \infty$ we always have

$$\mathfrak{I}_s = [-\mathfrak{d}_s, \infty)$$

and for $d \geq 2$ (finite) we have

$$\mathfrak{I}_s = [-\mathfrak{d}_s, \infty) \quad \textit{if and only if} \quad \mathfrak{p}_s(-\mathfrak{d}_s) \in \partial\mathcal{S}_d \cap \mathcal{S}_d.$$

Proof. We prove the statements first for finite $d \geq 2$ (part (a)) and then for $d = \infty$ (part (b)).

(a): Let $d \in \mathbb{N}$ with $d \geq 2$ and $s \in \mathcal{S}_d \setminus \{0\}$. By Definition 12.1.9 we have $0 \in \mathfrak{I}_s$, i.e., $\mathfrak{I}_s \neq \emptyset$.

(a-i) We show that if $c \in \mathfrak{I}_s$ then $[c, \infty) \subseteq \mathfrak{I}_s$. Let $c \in \mathfrak{I}_s$. Then by Definition 12.1.9, we have $s' := \mathfrak{p}_s(c) \in \mathcal{S}_d$ and, since $d \geq 2$ is finite, by Richter's Theorem 5.3.1 there exists an at most $k = \binom{n+d}{d}$-atomic representing measure

$$\mu = \sum_{i=1}^{k} c_i \cdot \delta_{x_i}$$

of s' with $c_i > 0$ and $x_i \in \mathbb{R}^n$. Moreover, by Theorem 12.1.8 for all $t > 0$ we have that $\mathfrak{p}_s(c+t)$ is represented by a non-negative Gaussian mixture. Hence, $\mathfrak{p}_s(c+t) \in \mathcal{S}_d$ for all $t > 0$ and $[c, \infty) \subseteq \mathfrak{I}_s$.

(a-ii) We show that $\mathfrak{d}_s := -\inf_{t \in \mathfrak{I}_s} t \leq \frac{s_{2e_1} + \cdots + s_{2e_n}}{2n \cdot s_0}$. We have

$$L_{\mathfrak{p}_s(t)}(x_1^2 + \cdots + x_n^2) = \sum_{i=1}^{n} \mathfrak{p}_{s,2e_i}(t) = \sum_{i=1}^{n} s_{2e_i} + 2ns_0 \cdot t \geq 0, \tag{28}$$

which implies the bound on t resp. its infimum $\mathfrak{d}_s$. By (a-i) we have $(-\mathfrak{d}_s, \infty) \subseteq \mathfrak{I}_s$.

(a-iii) "$\mathfrak{I}_s = [-\mathfrak{d}_s, \infty)$ if and only if $\mathfrak{p}_s(-\mathfrak{d}_s) \in \mathcal{S}_d$" follows directly from (a-ii) and Definition 12.1.9.

(b) We now prove the statements for $d = \infty$.

(b-i) We show $(-\mathfrak{d}_s, \infty) \subseteq \mathfrak{I}_s$. Let $k \in \mathbb{N}$, $s = (s_\alpha)_{\alpha \in \mathbb{N}_0^n} \in \mathcal{S}_\infty$, and denote by $s|_k \in \mathcal{S}_k$ the moment sequence truncated up to order k, i.e., $s|_k := (s_\alpha)_{|\alpha| \leq k}$. Then $(\mathfrak{d}_{s|_k})_{k \in \mathbb{N}}$ is a non-increasing sequence ≥ 0. Hence,

$$\mathfrak{d}_s = \lim_{k \to \infty} \mathfrak{d}_{s|_k} \geq 0$$

exists and $\mathfrak{d}_s \leq \mathfrak{d}_{s|k}$ implies

$$(-\mathfrak{d}_s, \infty) \subseteq \bigcap_{k \in \mathbb{N}} (-\mathfrak{d}_{s|_k}, \infty) \subseteq \mathfrak{I}_s.$$

(b-ii) We show $-\mathfrak{d}_s \in \mathfrak{I}_s$. Since $(-\mathfrak{d}_s, \infty) \subseteq \mathfrak{I}_s$ we have that $\mathfrak{p}_s(-\mathfrak{d}_s + \varepsilon) \in \mathcal{S}_\infty$ is a moment sequence for all $\varepsilon > 0$. Hence,

$$L_{\mathfrak{p}_s(-\mathfrak{d}_s + \varepsilon)}(p) \geq 0$$

for all $p \in \mathbb{R}[x_1,\dots,x_n]$ with $p \geq 0$ and $\varepsilon > 0$. Since $L_{\mathfrak{p}_s(-\mathfrak{d}_s+\varepsilon)}(p)$ is continuous in ε we have

$$L_{\mathfrak{p}_s(-\mathfrak{d}_s)}(p) = \lim_{\varepsilon\to 0} L_{\mathfrak{p}_s(-\mathfrak{d}_s+\varepsilon)}(p) \geq 0$$

for all $p \in \mathbb{R}[x_1,\dots,x_n]$ with $p \geq 0$. Therefore, $L_{\mathfrak{p}_s(-\mathfrak{d}_s)}$ has a representing measure and $\mathfrak{p}_s(-\mathfrak{d}_s) \in \mathcal{S}_\infty$, i.e., $-\mathfrak{d}_s \in \mathfrak{I}_s$. □

From Definition 12.1.5, we know that each component $\mathfrak{p}_{s,\alpha}(t) \in \mathbb{R}[t]$ depends only on s_β with $\beta \leq \alpha$. Hence, in the time evolution $L_{\mathfrak{p}_s(t)}(p_0)$ for $p_0 \in \mathbb{R}[x_1,\dots,x_n]$ we find that there is a $p_t(x) \in \mathbb{R}[x_1,\dots,x_n,t]$ such that

$$L_{\mathfrak{p}_s(t)}(p_0) = L_s(p_t)$$

for all $t \in \mathbb{R}$. This p_t can be found by rearranging

$$L_{\mathfrak{p}_s(t)}(p_0) = \sum_{\alpha:|\alpha|\leq \deg p_0} c_\alpha(t) \cdot s_\alpha$$

with $c_\alpha \in \mathbb{R}[t]$. Then p_t can be defined uniquely from the c_α's. Note that in the following definition and lemma, the polynomial p_t (just like $\mathfrak{p}_s(t)$) is defined for all $t \in \mathbb{R}$.

Definition 12.1.11. Let $p_0(x) = \sum_\alpha c_\alpha(0) \cdot x^\alpha \in \mathbb{R}[x_1,\dots,x_n]$. For any $\alpha \in \mathbb{N}_0^n$ with $|\alpha| \leq \deg(p_0)$ denote by $c_\alpha \in \mathbb{R}[t]$ the polynomial given by

$$L_{\mathfrak{p}_s(t)}(p_0) = \sum_{\alpha:|\alpha|\leq \deg p_0} c_\alpha(t) \cdot s_\alpha.$$

Then we define the polynomial $p_t(x) \in \mathbb{R}[x_1,\dots,x_n,t]$ as

$$p_t(x) := \sum_{\alpha:|\alpha|\leq \deg p_0} c_\alpha(t) \cdot x^\alpha$$

for all $t \in \mathbb{R}$.

Lemma 12.1.12. *Let $p_0 \in \mathbb{R}[x_1,\dots,x_n]$. Then the heat equation*

$$\begin{aligned} \partial_t u(x,t) &= \Delta u(x,t) \\ u(x,0) &= p_0(x) \end{aligned}$$

with initial data p_0 has the unique solution $u(x,t) = p_t(x) \in \mathbb{R}[x_1,\dots,x_n,t]$ for all $t \in \mathbb{R}$. Additionally, we have

$$\deg p_t = \deg p_0$$

for all $t \in \mathbb{R}$ where deg *is the degree in x.*

Proof. Let $v_0 \in C_c^\infty(\mathbb{R}^n) \subset \mathcal{S}(\mathbb{R}^n)$ and v be the unique solution of

$$\begin{aligned}\partial_t v(x,t) &= \Delta v(x,t)\\ v(x,0) &= v_0(x).\end{aligned}$$

Since $p_t \in \mathbb{R}[x_1,\dots,x_n]$ for all $t \in \mathbb{R}$ and $p_0 \in \mathbb{R}[x_1,\dots,x_n]$ is arbitrary, it is sufficient to prove $\partial_t p_t(x) = \Delta p_t(x)$ at $t = 0$. We have from Definition 12.1.11

$$\int_{\mathbb{R}^n} p_t(x) \cdot v_0(x)\ \mathrm{d}x = \int_{\mathbb{R}^n} p_0(x) \cdot v(x,t)\ \mathrm{d}x. \tag{29}$$

By differentiating (29) with respect to t we get

$$\begin{aligned}\partial_t \int_{\mathbb{R}^n} p_t(x) \cdot v_0(x)\ \mathrm{d}x\Big|_{t=0} &= \partial_t \int_{\mathbb{R}^n} p_0(x) \cdot v(x,t)\ \mathrm{d}x\Big|_{t=0}\\ &= \int_{\mathbb{R}^n} p_0(x) \cdot \partial_t v(x,t)\ \mathrm{d}x\Big|_{t=0}\\ &= \int_{\mathbb{R}^n} p_0(x) \cdot \Delta v(x,t)\ \mathrm{d}x\Big|_{t=0}\\ &= \int_{\mathbb{R}^n} (\Delta p_0(x)) \cdot v(x,t)\ \mathrm{d}x\Big|_{t=0}\\ &= \int_{\mathbb{R}^n} (\Delta p_0(x)) \cdot v_0(x)\ \mathrm{d}x.\end{aligned}$$

Since $v_0 \in C_c^\infty(\mathbb{R}^n)$ was arbitrary we have

$$\partial_t p_t(x)\big|_{t=0} = \Delta p_0(x)$$

which proves the statement. □

Remark 12.1.13. Lemma 12.1.12 can be also interpreted as follows. The unique solution of the heat equation is gained by convolution with Θ_t and we have the well-known relation

$$\begin{aligned}\int_{\mathbb{R}^n} g(x) \cdot (\Theta_t * f)(x)\ \mathrm{d}x &= \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} g(x) \cdot \Theta_t(x-y) \cdot f(y)\ \mathrm{d}x\ \mathrm{d}y\\ &= \int_{\mathbb{R}^n} (\Theta_t * g)(y) \cdot f(y)\ \mathrm{d}y\end{aligned}$$

for all $g \in \mathbb{R}[x_1,\dots,x_n]$ and f a Schwartz function on $\mathbb{R}^n$. For a measure μ_0 which has finite moments we define $\mu_t = \Theta_t * \mu_0$ in the same manner

$$\int_{\mathbb{R}^n} f(x)\ \mathrm{d}\mu_t(x) = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} f(x) \cdot \Theta_t(x-y)\ \mathrm{d}\mu_0(y) \tag{30a}$$

which is by Fubini

$$= \int_{\mathbb{R}^n} (\Theta_t * f)(y)\ \mathrm{d}\mu_0(y) \tag{30b}$$

and μ_t solves the heat equation. Together, (30a) and (30b) can be used to provide an alternative proof of Lemma 12.1.12. ◦

Example 12.1.14. Let $n = 1$ and $s = (s_k(0))_{k=0}^2$ be a real sequence. We have

$$\mathfrak{p}_{s,0}(t) = s_0(0), \quad \mathfrak{p}_{s,1}(t) = s_1(0), \quad \text{and} \quad \mathfrak{p}_{s,2}(t) = s_2(0) + 2s_0(0)\cdot t;$$

see (27). For $p_0(x) = c_0(0) + c_1(0)\cdot x + c_2(0)\cdot x^2$ we therefore have

$$\begin{aligned} L_{\mathfrak{p}_s(t)}(p_0) &= c_0(0)\cdot s_0(t) + c_1(0)\cdot s_1(t) + c_2(0)\cdot s_2(t) \\ &= c_0(0)\cdot s_0(0) + c_1(0)\cdot s_1(0) + c_2(0)\cdot [s_2(0) + 2s_0(0)\cdot t] \\ &= [c_0(0) + 2c_2(0)\cdot t]\cdot s_0(0) + c_1(0)\cdot s_1(0) + c_2(0)\cdot s_2(0) \\ &= c_0(t)\cdot s_0(0) + c_1(t)\cdot s_1(0) + c_2(t)\cdot s_2(0) \\ &= L_s(p_t), \end{aligned}$$

i.e.,

$$\begin{aligned} p_t(x) &= [c_0(0) + 2c_2(0)\cdot t] + c_1(0)\cdot x + c_2(0)\cdot x^2, \\ \partial_t p_t(x) &= 2c_2(0), \end{aligned}$$

and

$$\partial_x^2 p_t(x) = 2c_2(0)$$

for all $t \in \mathbb{R}$. ∘

Theorem 12.1.15. *Let $n \in \mathbb{N}$ and $s \in \mathcal{S}_\infty$ be an indeterminate moment sequence. Then $\mathfrak{p}_s(t)$ is indeterminate for all $t \in [0,\infty)$.*

Proof. First we prove that $\mathfrak{p}_s(t)$ is indeterminate for all $t \in [0,\varepsilon)$ for some $\varepsilon > 0$:

Since s is indeterminate, and it has at least two distinct representing measures μ_0 and $\tilde{\mu}_0$. Since μ_0 and $\tilde{\mu}_0$ are distinct there exists a measurable $A \subset \mathbb{R}^n$ such that

$$\int_{\mathbb{R}^n} \chi_A(x)\ \mathrm{d}\mu_0(x) \neq \int_{\mathbb{R}^n} \chi_A(x)\ \mathrm{d}\tilde{\mu}_0 \tag{31}$$

where χ_A is the characteristic function of A. As the representing measure is Radon, we can assume without loss of generality that A is compact. For the time-dependent measures μ_t and $\tilde{\mu}_t$ we find from (30) that

$$\int_{\mathbb{R}^n} \chi_A(x)\ \mathrm{d}\mu_t(x) = \int_{\mathbb{R}^n} (\Theta_t * \chi_A)\ \mathrm{d}\mu_0(x) \tag{32}$$

$$\int_{\mathbb{R}^n} \chi_A(x)\ \mathrm{d}\tilde{\mu}_t(x) = \int_{\mathbb{R}^n} (\Theta_t * \chi_A)\ \mathrm{d}\tilde{\mu}_0(x) \tag{33}$$

Both (32) and (33) continuously depend on $t \geq 0$ and since for $t = 0$ we have (31) there exists an $\varepsilon > 0$ such that (32) $\neq$ (33) for all $t \in [0,\varepsilon)$, i.e., μ_t and $\tilde{\mu}_t$ are two distinct representing measures of $\mathfrak{p}_s(t)$ and hence $\mathfrak{p}_s(t)$ is indeterminate for all $t \in [0,\varepsilon)$.

Now we show that for $t = \varepsilon/2$ there are C^∞-functions $f_{\varepsilon/2}$ and $\tilde{f}_{\varepsilon/2}$ such that

$$\mathrm{d}\mu_{\varepsilon/2}(x) = f_{\varepsilon/2}(x)\ \mathrm{d}x \qquad \text{and} \qquad \mathrm{d}\tilde{\mu}_{\varepsilon/2}(x) = \tilde{f}_{\varepsilon/2}(x)\ \mathrm{d}x.$$

It is sufficient to show this for μ_0:

Since $s_0 = \int_{\mathbb{R}^n} 1 \; \mathrm{d}\mu_0 < \infty$, we have $\mu_0(A) < \infty$ for all Borel-measurable sets $A \in \mathcal{B}(\mathbb{R}^n)$. Let $\nu := e^{-x^2} \cdot \lambda$, where λ is the Lebesgue measure on $\mathbb{R}^n$. Then μ_0 and ν are finite measures. Hence, by the Lebesgue decomposition [Bog07, Thm. 3.2.3] there exists a ν-integrable function g and a measure ρ such that

$$\mu_0 = g \cdot \nu + \rho,$$

ρ is singular with respect to ν, i.e., there exists $A \in \mathcal{B}(\mathbb{R}^n)$ with $\nu(A) = 0$ but $\rho(A) > 0$.

We show that $\Theta_t * \rho$ for $t > 0$ is no longer singular with respect to ν: Let $A \in \mathcal{B}(\mathbb{R}^n)$ with $\nu(A) = 0$. Then also the Lebesgue measure fulfills $\lambda(A) = 0$. Since

$$\chi_A(x) := \begin{cases} 1 & \text{for } x \in A, \\ 0 & \text{else} \end{cases} \qquad \Rightarrow \qquad \Theta_t * \chi_A = 0 \text{ for all } t > 0,$$

we have

$$\begin{aligned}(\Theta_t * \rho)(A) &= \int_{\mathbb{R}^n} \chi_A(x) \; \mathrm{d}(\Theta_t * \rho)(x) = \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \chi_A(x) \cdot \Theta_t(x-y) \; \mathrm{d}\rho(y) \; \mathrm{d}x \\ &= \int_{\mathbb{R}^n} (\Theta_t * \chi_A)(y) \; \mathrm{d}\rho(y) = \int_{\mathbb{R}^n} 0 \; \mathrm{d}\rho(y) = 0.\end{aligned}$$

Hence, for $t = \varepsilon/4$ we have that $\Theta_{\varepsilon/4} * \rho = h \cdot \nu$ for a ν-integrable function h. In summary for $t = \varepsilon/2$ we have

$$\begin{aligned}\mu_{\varepsilon/2} &= \Theta_{\varepsilon/2} * \mu_0 \\ &= \Theta_{\varepsilon/2} * (g \cdot \nu) + \Theta_{\varepsilon/2} * \rho \\ &= [\Theta_{\varepsilon/2} * (g \cdot e^{-x^2})] \cdot \lambda + \Theta_{\varepsilon/4} * (h \cdot \nu) \\ &= [\Theta_{\varepsilon/2} * (g \cdot e^{-x^2})] \cdot \lambda + [\Theta_{\varepsilon/4} * (h \cdot e^{-x^2})] \cdot \lambda \\ &= [\Theta_{\varepsilon/2} * (g \cdot e^{-x^2}) + \Theta_{\varepsilon/4} * (h \cdot e^{-x^2})] \cdot \lambda \\ &= f_{\varepsilon/2} \cdot \lambda\end{aligned}$$

with $f_{\varepsilon/2}$ a C^∞-function. In the same way we get $\tilde{\mu}_{\varepsilon/2} = \tilde{f}_{\varepsilon/2} \cdot \lambda$ for a C^∞-function $\tilde{f}_{\varepsilon/2}$. We already showed that $\mu_t \neq \tilde{\mu}_t$ for all $t \in [0, \varepsilon)$, i.e., for $t = \varepsilon/2$ we get $f_{\varepsilon/2} \neq \tilde{f}_{\varepsilon/2}$.

Since the heat equation has the backwards uniqueness, see e.g. [Eva10, Ch. 2.3], we have

$$\Theta_t * f_{\varepsilon/2} \neq \Theta_t * \tilde{f}_{\varepsilon/2}$$

for all $t \geq 0$, i.e., $\mu_t \neq \tilde{\mu}_t$ for all $t \geq 0$. Therefore, $\mathfrak{p}_t(s)$ is an indeterminate moment sequence for all $t \geq 0$. □

Corollary 12.1.16. *Let $n \in \mathbb{N}$ and $s \in \mathcal{S}_\infty$ be a determinate moment sequence. Then $\mathfrak{p}_s(t)$ is a determinate moment sequence for all $t \in \mathfrak{I}_s \cap (-\infty, 0]$.*

Proof. Assume $\mathfrak{p}_s(t)$ is an indeterminate moment sequence for some $t < 0$. Then, by Theorem 12.1.15, we also know that $s = \mathfrak{p}_s(0)$ is indeterminate, a contradiction. □

In Theorem 12.1.15 we also proved the following result.

Theorem 12.1.17. *Let $n \in \mathbb{N}$, $s \in \mathcal{S}_\infty$, and λ be the Lebesgue measure on $\mathbb{R}^n$. Then there exists a family $\{f_t\}_{t>0} \subset \mathcal{S}(\mathbb{R}^n)$ with*

$$f_{t_1+t_2} = \Theta_{t_1} * f_{t_2} = \Theta_{t_2} * f_{t_1}$$

for all $t_1, t_2 > 0$ such that $\mathfrak{p}_s(t)$ is represented by $f_t \cdot \lambda$ for all $t > 0$, i.e.,

$$\mathfrak{p}_{s,\alpha}(t) = \int_{\mathbb{R}^n} x^\alpha \cdot f_t(x)\ \mathrm{d}x \qquad \text{for all } \alpha \in \mathbb{N}_0^n \text{ and } t > 0. \tag{34}$$

Proof. In Theorem 12.1.15 we already established that $\{f_t\}_{t>0}$ is a family of C^∞-functions such that (34) holds. It remains to show that

$$\|x^\alpha \cdot \partial^\beta f_t(x)\|_\infty < \infty \tag{35}$$

for all $\alpha, \beta \in \mathbb{N}_0^n$ and $t > 0$ to have $\{f_t\}_{t>0} \subset \mathcal{S}(\mathbb{R}^n)$.

Let $t > 0$. Since $\int_{\mathbb{R}^n} x^\alpha \cdot f_t(x)\ \mathrm{d}x$ is finite by the definition of the Lebesgue integral we have $\int_{\mathbb{R}^n} |x^\alpha \cdot f_t(x)|\ \mathrm{d}x < \infty$ for all $\alpha \in \mathbb{N}_0^n$, i.e., $\lim_{|x|\to\infty} x^\alpha \cdot f_t(x) = 0$ for all $\alpha \in \mathbb{N}_0^n$. Therefore, we can use partial integration to get

$$\int_{\mathbb{R}^n} x^\alpha \cdot \partial^\beta f_t(x)\ \mathrm{d}x = (-1)^{|\alpha|} \cdot \int_{\mathbb{R}^n} (\partial^\beta x^\alpha) \cdot f_t(x)\ \mathrm{d}x. \tag{36}$$

It is therefore sufficient to show (35) for $\beta = 0$.

Since $g_\alpha(x) = x^\alpha \cdot f_t(x)$ is continuous and $g_\alpha \in L^1(\mathbb{R}^n)$, we also have $g_\alpha \in L^2(\mathbb{R}^n)$. By (36) we also have that $\partial^\beta g_\alpha \in L^2(\mathbb{R}^n)$ for all $\beta \in \mathbb{N}_0^n$. Hence, $g_\alpha \in H^\infty(\mathbb{R}^n)$ for all $\alpha \in \mathbb{N}_0^n$ and by the Sobolev Imbedding Theorem we have (35) for all $\alpha, \beta \in \mathbb{N}_0^n$. $\square$

12.2. The Heat Equation acting on Polynomials

The following are the explicit time-dependent polynomials for the 1-dimensional heat equation in Lemma 12.1.12.

Definition 12.2.1. Let $d \in \mathbb{N}_0$. We define $\mathfrak{p}_{2d}, \mathfrak{p}_{2d+1} \in \mathbb{R}[x,t]$ by

$$\mathfrak{p}_{2d}(x,t) := \sum_{j=0}^{d} \frac{(2d)!}{(2d-2j)! \cdot j!} \cdot t^j \cdot x^{2d-2j}$$

and

$$\mathfrak{p}_{2d+1}(x,t) := \sum_{j=0}^{d} \frac{(2d+1)!}{(2d+1-2j)! \cdot j!} \cdot t^j \cdot x^{2d+1-2j}.$$

Example 12.2.2. We have

$$\begin{aligned}
\mathfrak{p}_0(x,t) &= 1\\
\mathfrak{p}_1(x,t) &= x\\
\mathfrak{p}_2(x,t) &= 2t + x^2\\
\mathfrak{p}_3(x,t) &= 6tx + x^3\\
\mathfrak{p}_4(x,t) &= 12t^2 + 12tx^2 + x^4\\
\mathfrak{p}_5(x,t) &= 60t^2x + 20tx^3 + x^5\\
\mathfrak{p}_6(x,t) &= 120t^3 + 180t^2x^2 + 30tx^4 + x^6\\
&\vdots
\end{aligned}$$

$\circ$

Straightforward calculations show that $\mathfrak{p}_k$, $k \in \mathbb{N}_0$, solve the initial value heat equation

$$\begin{aligned}
\partial_t \mathfrak{p}_k(x,t) &= \partial_x^2 \mathfrak{p}_k(x,t)\\
\mathfrak{p}_k(x,0) &= x^k.
\end{aligned} \tag{37}$$

Hence, by linearity of the heat equation we have the following extension of Definition 12.2.1 and the observation (37), the explicit version of Lemma 12.1.12.

Theorem 12.2.3. *Let $d \in \mathbb{N}_0$, $n \in \mathbb{N}$, and $f_0(x) = \sum_{\alpha\in\mathbb{N}_0^n} c_\alpha \cdot x^\alpha \in \mathbb{R}[x_1,\dots,x_n]$. Then*

$$\mathfrak{p}_{f_0}(x,t) := \sum_{\alpha\in\mathbb{N}_0^n} c_\alpha \cdot \mathfrak{p}_{\alpha_1}(x_1,t)\cdots \mathfrak{p}_{\alpha_n}(x_n,t) \quad \in \mathbb{R}[x_1,\dots,x_n,t] \tag{38}$$

with $\alpha = (\alpha_1,\dots,\alpha_n) \in \mathbb{N}_0^n$ solves the initial value heat equation

$$\begin{aligned}
\partial_t f(x,t) &= \Delta f(x,t)\\
f(x,0) &= f_0(x).
\end{aligned}$$

Proof. By linearity of the Laplace operator Δ it is sufficient to look at $f_0(x) = x^\alpha$ for $\alpha \in \mathbb{N}_0^n$. By (37) we already have $\partial_t \mathfrak{p}_{\alpha_i}(x_i) = \partial_i^2 \mathfrak{p}_{\alpha_i}(x)$ and hence

$$\begin{aligned}
\partial_t \mathfrak{p}_{f_0}(x,t) &= \partial_t[\mathfrak{p}_{\alpha_1}(x_1,t)\cdots\mathfrak{p}_{\alpha_n}(x_n,t)]\\
&= [\partial_t \mathfrak{p}_{\alpha_1}(x_1,t)]\cdot \mathfrak{p}_{\alpha_2}(x_2,t)\cdots\mathfrak{p}_{\alpha_n}(x_n,t)\\
&\qquad + \dots + \mathfrak{p}_{\alpha_1}(x_1,t)\cdots\mathfrak{p}_{\alpha_{n-1}}(x_{n-1},t)\cdot[\partial_t\mathfrak{p}_{\alpha_n}(x_n,t)]\\
&= [\partial_1^2 \mathfrak{p}_{\alpha_1}(x_1,t)]\cdot \mathfrak{p}_{\alpha_2}(x_2,t)\cdots\mathfrak{p}_{\alpha_n}(x_n,t)\\
&\qquad + \dots + \mathfrak{p}_{\alpha_1}(x_1,t)\cdots\mathfrak{p}_{\alpha_{n-1}}(x_{n-1},t)\cdot[\partial_n^2\mathfrak{p}_{\alpha_n}(x_n,t)]\\
&= \Delta \mathfrak{p}_{f_0}(x,t).
\end{aligned}$$

$\square$

Example 12.2.4 (Motzkin polynomial [Mot67])**.** Let

$$f_{\mathrm{Motz}}(x,y) = 1 - 3x^2y^2 + x^4y^2 + x^2y^4 \in \mathbb{R}[x,y]$$

be the Motzkin polynomial. Then by Definition 12.2.1 (resp. Example 12.2.2) we have the substitutions

$$\begin{aligned} x^2 &\mapsto 2t + x^2, & x^4 &\mapsto 12t^2 + 12tx^2 + x^4,\\ y^2 &\mapsto 2t + y^2, & y^4 &\mapsto 12t^2 + 12ty^2 + y^4 \end{aligned}$$

and get

$$\begin{aligned} \mathfrak{p}_{\mathrm{Motz}}(x,y,t) &= 1 - 3(2t+x^2)(2t+y^2) + (12t^2 + 12tx^2 + x^4)(2t+y^2)\\ &\quad + (2t+x^2)(12t^2+12ty^2+y^4)\\ &= 1 - 12t^2 + 48t^3 + 6t(-1+6t)(x^2+y^2) + (-3+24t)x^2y^2\\ &\quad + 2t(x^4+y^4) + x^4y^2 + x^2y^4 \end{aligned}$$

for all $t \in \mathbb{R}$. ◦

Theorem 12.2.5. *Let $d \in \mathbb{N}_0$ and ρ be a kernel such that $\int_{\mathbb{R}^n} y^\alpha \cdot \rho(y)\ \mathrm{d}y$ is finite for all $\alpha \in \mathbb{N}_0^n$ with $|\alpha| \leq d$, then*

$$\cdot * \rho : \mathbb{R}[x_1,\dots,x_n]_{\leq d} \to \mathbb{R}[x_1,\dots,x_n]_{\leq d}.$$

Proof. Let $p \in \mathbb{R}[x_1,\dots,x_n]_{\leq d}$. Then from

$$(p*\rho)(x) = \int_{\mathbb{R}^n} p(x-y)\cdot\rho(y)\ \mathrm{d}y$$

and expanding $p(x-y)$ in the right side gives the assertion including the degree bound $\deg(p*\rho) \leq d$. □

Corollary 12.2.6. *Let $n \in \mathbb{N}$, $d \in \mathbb{N}_0$, and $f_0 \in \mathrm{Pos}(n,d)$. Then*

$$\mathfrak{p}_{f_0}(\,\cdot\,,t) \in \mathrm{Pos}(n,d)$$

for all $t \geq 0$. Especially, if $f_0 \neq 0$ then $\mathfrak{p}_{f_0}(\,\cdot\,,t) > 0$ on $\mathbb{R}^n$ for all $t > 0$.

Corollary 12.2.7. *Let $n \in \mathbb{N}$ and $f_0 \in \mathbb{R}[x_1,\dots,x_n]$. Assume there exist $t > 0$ and a point $\xi \in \mathbb{R}^n$ such that $\mathfrak{p}_{f_0}(\xi,t) < 0$. Then $f_0 \notin \mathrm{Pos}(n,d)$.*

Example 12.2.8 (Motzkin polynomial, Example 12.2.4 continued)**.** We have

$$\begin{aligned} \mathfrak{p}_{\mathrm{Motz}}(x,y,1) &= 37\cdot\left(1 - \frac{11}{148}x^2 - \frac{11}{148}y^2\right)^2 + \frac{71}{2}\cdot\left(x - \frac{4}{71}xy^2\right)^2\\ &\quad + \frac{71}{2}\cdot\left(y - \frac{4}{71}x^2y\right)^2 + \frac{57}{2}x^2y^2 + \frac{1063}{592}\cdot\left(x^2 + \frac{27}{1063}y^2\right)^2\\ &\quad + \frac{3815}{2126}\cdot y^4 + \frac{63}{71}\cdot x^4y^2 + \frac{63}{71}x^2y^4 \quad \in \mathrm{SOS}(2,6) \end{aligned}$$

i.e., $\mathfrak{p}_{\mathrm{Motz}}(\,\cdot\,,1)$ is by Corollary 12.2.6 not just non-negative, but in fact a sum of squares. This relation can easily be obtained e.g. by the use of Macaulay2 [GS] and the SumsOfSquares package [CKP20]. In fact, additional calculations indicate that

$$\mathfrak{p}_{\mathrm{Motz}}(\,\cdot\,,t) \in \begin{cases} \mathrm{Pos}(2,6)\setminus\mathrm{SOS}(2,6) & \text{for } t\in[0,T_{\mathrm{Motzkin}}), \text{ and}\\ \mathrm{SOS}(2,6) & \text{for } t\in[T_{\mathrm{Motzkin}},\infty),\end{cases}$$

with

$$\frac{31\,998}{1\,000\,000} \quad<\quad T_{\mathrm{Motzkin}} \quad<\quad \frac{31\,999}{1\,000\,000}.$$

The choice of the intervals $[0,T_{\mathrm{Motzkin}})$ and $[T_{\mathrm{Motzkin}},\infty)$ is clear since $\mathrm{SOS}(2,6)$ is closed and $\mathfrak{p}_{f_0}(\,\cdot\,,t)$ continuous in t, i.e., $\mathfrak{p}_{\mathrm{Motz}}(\,\cdot\,,T_{\mathrm{Motzkin}})\in\mathrm{SOS}(2,6)$. ∘

Theorem 12.2.9. *Let $\rho\geq 0$ be a kernel such that $\int_{\mathbb{R}^n} y^\alpha\cdot\rho(y)\;\mathrm{d}y$ is finite for all $\alpha\in\mathbb{N}_0^n$ with $|\alpha|\leq d$, then*

$$\cdot*\rho:\mathrm{SOS}(n,d)\to\mathrm{SOS}(n,d).$$

Proof. Let $p\in\mathrm{SOS}(n,d)$, i.e., there exists a symmetric $Q\in\mathbb{R}^{N\times N}$ with $N=\binom{n+d}{d}$ such that $p(x)=(x^\alpha)_\alpha^T\cdot Q\cdot(x^\alpha)_\alpha$ where $(x^\alpha)_\alpha$ is the vector of all monomials x^α with $|\alpha|\leq d$. We then have

$$\begin{aligned}(p*\rho)(x) &= \int_{\mathbb{R}^n} p(x-y)\cdot\rho(y)\;\mathrm{d}y\\ &= \int_{\mathbb{R}^n} ((x-y)^\alpha)_\alpha^T\cdot Q\cdot((x-y)^\alpha)_\alpha\cdot\rho(y)\;\mathrm{d}y\end{aligned}$$

and by Richter's Theorem 5.3.1 we can replace $\rho(y)\;\mathrm{d}y$ by a finitely atomic representing measure $\mu=\sum_{i=1}^k c_i\cdot\delta_{y_i}$ with $c_i>0$ and get

$$=\sum_{i=1}^k c_i\cdot((x-y_i)^\alpha)_\alpha^T\cdot Q\cdot((x-y_i)^\alpha)_\alpha\in\mathrm{SOS}(n,d).$$ □

Theorem 12.2.10. *Let $\rho\geq 0$ be a kernel such that $\int_{\mathbb{R}^n} y^\alpha\cdot\rho(y)\;\mathrm{d}y$ is finite for all $\alpha\in\mathbb{N}_0^n$, then*

$$\cdot*\rho:\mathrm{War}(n,d)\to\mathrm{War}(n,d).$$

Proof. Let $p\in\mathrm{War}(n,d)$, i.e., $p(x)=\sum_{i=1}^k(a_i\cdot x)^d$. Then

$$\begin{aligned}(p*\rho)(x) &= \int_{\mathbb{R}^n} p(x-y)\cdot\rho(y)\;\mathrm{d}y\\ &= \int_{\mathbb{R}^n}\sum_{i=1}^k(a_i\cdot(x-y))^d\cdot\rho(y)\;\mathrm{d}y\end{aligned}$$

and by Richter's Theorem 5.3.1 we can replace $\rho(y)\;\mathrm{d}y$ by a finitely atomic representing measure $\mu=\sum_{j=1}^l c_j\cdot\delta_{y_j}$ with $c_j>0$ and get

$$=\sum_{j=1}^l\sum_{i=1}^k c_j\cdot(a_i\cdot(x-y_j))^d\in\mathrm{War}(n,d).$$ □

In Examples 4.3.3 we listed several non-negative polynomials which are not sums of squares. We want to investigate, what happens to them under the heat equation.

Example 12.2.11 (Robinson polynomial [Rob69])**.** Let

$$f_{\mathrm{Robinson}}(x,y) = 1 - x^2 - y^2 - x^4 + 3x^2y^2 - y^4 + x^6 - x^4y^2 - x^2y^4 + y^6$$

be the Robinson polynomial, i.e., $f_{\mathrm{Robinson}} \in \mathrm{Pos}(2,6) \setminus \mathrm{SOS}(2,6)$. Then by a direct calculation using Macaulay2 with the SumsOfSquares package similar to the Motzkin polynomial we find $\mathfrak{p}_{\mathrm{Robinson}}(\,\cdot\,,1) \in \mathrm{SOS}(2,6)$ and by Theorem 12.2.9 we have

$$\mathfrak{p}_{\mathrm{Robinson}}(\,\cdot\,,t) \in \begin{cases} \mathrm{Pos}(2,6)\setminus\mathrm{SOS}(2,6) & \text{for } t \in [0,T_{\mathrm{Robinson}}), \text{ and}\\ \mathrm{SOS}(2,6) & \text{for } t \in [T_{\mathrm{Robinson}},\infty),\end{cases}$$

with

$$\frac{20\,946}{1\,000\,000} \quad < \quad T_{\mathrm{Robinson}} \quad < \quad \frac{20\,947}{1\,000\,000}. \qquad \circ$$

Example 12.2.12 (Choi–Lam polynomial [CL77])**.** Let

$$f_{\mathrm{Choi-Lam}}(x,y,z) = 1 - 4xyz + x^2y^2 + x^2z^2 + y^2z^2$$

be the Choi–Lam polynomial, i.e., $f_{\mathrm{Choi-Lam}} \in \mathrm{Pos}(3,4)\setminus\mathrm{SOS}(3,4)$. We have

$$\begin{aligned}\mathfrak{p}_{\mathrm{Choi-Lam}}(x,y,z,t) &= 1 - 4xyz + (2t+x^2)(2t+y^2) + (2t+x^2)(2t+z^2)\\ &\quad + (2t+y^2)(2t+z^2)\\ &= 1 + 12t^2 - 4xyz + 4t(x^2+y^2+z^2) + x^2y^2 + x^2z^2 + y^2z^2\\ &= f_{\mathrm{Choi-Lam}}(x,y,z) + 12t^2 + 4t(x^2+y^2+z^2)\end{aligned}$$

for all $t \in \mathbb{R}$ and for $t = \frac{1}{9}$ we have

$$\mathfrak{p}_{\mathrm{Choi-Lam}}(x,y,z,1/9) = \frac{31}{27} + \left(xy - \frac{2}{3}z\right)^2 + \left(xz - \frac{2}{3}y\right)^2 + \left(yz - \frac{2}{3}x\right)^2. \qquad \circ$$

We have

$$\mathfrak{p}_{\mathrm{Choi-Lam}}(\,\cdot\,,t) \in \begin{cases} \mathrm{Pos}(3,4)\setminus\mathrm{SOS}(3,4) & \text{for } t \in [0,T_{\mathrm{Choi-Lam}}), \text{ and}\\ \mathrm{SOS}(3,4) & \text{for } t \in [T_{\mathrm{Choi-Lam}},\infty),\end{cases}$$

with

$$\frac{1}{9} - 7\cdot 10^{-9} \quad < \quad T_{\mathrm{Choi-Lam}} \quad < \quad \frac{1}{9} - 6\cdot 10^{-9} \qquad \circ$$

Example 12.2.13 (Schmüdgen polynomial [Sch79])**.** The polynomial

$$\begin{aligned}f_{\mathrm{Schm}}(x,y) &= (y^2-x^2)x(x+2)[x(x-2)+2(y^2-4)]\\ &\quad + 200[(x^3-4x)^2 + (y^3-4y)^2] \quad \in \mathrm{Pos}(2,6)\setminus\mathrm{SOS}(2,6)\end{aligned}$$

is the Schmüdgen polynomial and we find $\mathfrak{p}_{\mathrm{Schm}}(\,\cdot\,,1) \in \mathrm{SOS}(2,6)$. In fact, Macaulay2 calculations with the SumsOfSquares package and Theorem 12.2.9 shows that $\mathfrak{p}_{\mathrm{Schm}}(\,\cdot\,,t) \in \mathrm{SOS}(2,6)$ for all $t \geq 2\cdot 10^{-4}$. $\circ$

Example 12.2.14 (Berg–Christensen–Jensen polynomial [BCJ79])**.** The Berg–Christensen–Jensen polynomial

$$f_{\mathrm{BCJ}}(x,y) = 1 - x^2y^2 + x^4y^2 + x^2y^4 \in \mathrm{Pos}(2,6) \setminus \mathrm{SOS}(2,6)$$

is connected to the Motzkin polynomial f_{Motzkin} (Example 12.2.4) by

$$f_{\mathrm{BCJ}}(x,y) = f_{\mathrm{Motzkin}}(x,y) + 2x^2y^2$$

and hence from Theorem 12.2.9 we see that

$$\mathfrak{p}_{\mathrm{BCJ}}(\,\cdot\,,t) \in \mathrm{SOS}(2,6)$$

for all $t \geq \frac{1}{6}$. ∘

Example 12.2.15 (Harris polynomial [Har99, $R_{2,0}$ in Lem. 5.1 and 6.8])**.** Let

$$\begin{aligned} f_{\mathrm{Har}}(x,y) = \quad & 16x^{10} - 36x^8y^2 + 20x^6y^4 + 20x^4y^6 - 36x^2y^8 + 16y^{10} \\ & - 36x^8 \;\; + 57x^6y^2 - 38x^4y^4 + 57x^2y^6 - 36y^8 \\ & + 20x^6 \;\; - 38x^4y^2 - 38x^2y^4 + 20y^6 \\ & + 20x^4 \;\; + 57x^2y^2 + 20y^4 \\ & - 36x^2 \;\; - 36y^2 \\ & + 16 \end{aligned}$$

be the Harris polynomial, i.e., $f_{\mathrm{Har}} = R_{2,0} \in \mathrm{Pos}(2,10)\backslash\mathrm{SOS}(2,10)$. With Example 12.2.2,

$$\mathfrak{p}_8(x,t) = 1680t^4 + 3360t^3x^2 + 840t^2x^4 + 56tx^6 + x^8,$$

and

$$\mathfrak{p}_{10}(x,t) = 30240t^5 + 75600t^4x^2 + 25200t^3x^4 + 2520t^2x^6 + 90tx^8 + x^{10}$$

we calculate $\mathfrak{p}_{\mathrm{Har}}$ and find $\mathfrak{p}_{\mathrm{Har}}(\,\cdot\,,1) \in \mathrm{SOS}(2,10)$. In fact, Macaulay2 calculations and Theorem 12.2.9 show that $\mathfrak{p}_{\mathrm{Har}}(\,\cdot\,,t) \in \mathrm{SOS}(2,10)$ for all $t \geq 8 \cdot 10^{-4}$. ∘

Lemma 12.2.16. *Let $n, d \in \mathbb{N}$ and*

$$f(x) = \sum_{|\alpha|\leq 2d} a_\alpha \cdot x^\alpha \in \mathrm{Pos}(n,2d)$$

such that

$$f_{2d}(x) := \sum_{|\alpha|=2d} a_\alpha \cdot x^\alpha \notin \mathrm{SOS}(n,2d),$$

then $\mathfrak{p}_f(\,\cdot\,,t) \in \mathrm{Pos}(n,2d) \setminus \mathrm{SOS}(n,2d)$ for all $t \geq 0$.

Proof. Assume there is a $t \geq 0$ such that

$$\mathfrak{p}_f(x,t) = \sum_{i=1}^{k} \left(\sum_{|\alpha|\leq d} c_{i,\alpha}(t) \cdot x^\alpha \right)^2 = \sum_{|\alpha|\leq 2d} a_\alpha(t) \cdot x^\alpha \in \mathrm{SOS}(n,2d).$$

Since by Definition 12.2.1 we have $a_\alpha(t) = a_\alpha(0)$ for all $\alpha \in \mathbb{N}_0^n$ with $|\alpha| = 2d$ the sum of squares decomposition of $\mathfrak{p}_f(\,\cdot\,,t)$ gives

$$f_{2d}(x) = \sum_{i=1}^{k} \left(\sum_{|\alpha|=d} c_{i,\alpha}(t) \cdot x^\alpha \right)^2 \in \mathrm{SOS}(n,2d)$$

which contradicts the assumption $f_{2d} \notin \mathrm{SOS}(n,2d)$. □

Example 12.2.17. Let $f(x,y,z) = z^6 - 3x^2y^2z^2 + x^4y^2 + x^2y^4 \in \mathrm{Pos}(3,6) \setminus \mathrm{SOS}(3,6)$ be the homogeneous Motzkin polynomial. Then $\mathfrak{p}_f(\,\cdot\,,t) \in \mathrm{Pos}(3,6) \setminus \mathrm{SOS}(3,6)$ for all $t \geq 0$. ○

Lemma 12.2.18. *Let $n \in \mathbb{N}$ and $f \in \mathrm{Pos}(\mathbb{R}^n)$ with $\deg f = 2d$ for some $d \in \mathbb{N}$. Then*

$$\lim_{t\to\infty} \frac{\mathfrak{p}_f(x,t)}{t^d} = c > 0.$$

Proof. Let $k \in \mathbb{N}$. It is easy to see that $\Delta^k g$ is constant on $\mathbb{R}^n$ for all $g \in \mathbb{R}[x_1,\dots,x_n]_{\leq 2k}$ and even equal to zero for all $g \in \mathbb{R}[x_1,\dots,x_n]_{\leq 2k-1}$.

Since $f \in \mathrm{Pos}(\mathbb{R}^n)$ with $\deg f = 2d$ we have that the homogeneous part f_{2d} of f of degree $2d$ is non-zero and non-negative on $\mathbb{R}^n$. Then by Theorem 8.3.5 we have $\partial_t^d \mathfrak{p}_f(x,t) = \Delta^d \mathfrak{p}_f(x,t) = \Delta^d f_{2d}(x) = c > 0$ which proves the statement. □

Open Problem 12.2.19. *Let $f \in \mathrm{Pos}(\mathbb{R}^2) \setminus \mathrm{SOS}(\mathbb{R}^2)$. Is it true that there always is a $T = T(f) > 0$ such that $\mathfrak{p}_f(x,t) \in \mathrm{SOS}(\mathbb{R}^2)$ for all $t \geq T$?*

Open Problem 12.2.20. *Let $f \in \mathrm{Pos}(\mathbb{R}^3) \setminus \mathrm{SOS}(\mathbb{R}^3)$ with $\deg f \leq 4$. Is it true that there always is a $T = T(f) > 0$ such that $\mathfrak{p}_f(x,t) \in \mathrm{SOS}(\mathbb{R}^3)$ for all $t \geq T$?*

12.3. Burgers' Equation

Theorem 12.3.1. *Let $u_0 \in \mathcal{S}(\mathbb{R},\mathbb{R})$. Then there exist maximal $T_1, T_2 > 0$ such that Burgers' equation*

$$\begin{aligned} \partial_t u &= -u \cdot \partial_x u \\ u(\,\cdot\,,0) &= u_0 \end{aligned} \tag{39}$$

has a unique classical solution $u \in C^\infty((-T_1,T_2),\mathcal{S}(\mathbb{R},\mathbb{R}))$. $(-T_1,T_2)$ is the maximal interval such that $u \in C((-T_1,T_2),C_b^\infty(\mathbb{R},\mathbb{R}))$.

Theorem 12.3.2. *Let $u_0 \in \mathcal{S}(\mathbb{R},\mathbb{R})$. Then for all $p \in \mathbb{N}$ and $k \in \mathbb{N}_0$ the time-dependent moments*

$$s_{k,p}(t) := \int_{\mathbb{R}} x^k \cdot u(x,t)^p \,\mathrm{d}x$$

of the solution u of Burgers' equation (39) are

$$s_{k,p}(t) = \sum_{i=0}^{k} \frac{s_{k-i,p+i}(0)}{i!} \cdot t^i \cdot \prod_{j=0}^{i-1} \frac{(p+j)\cdot(k-j)}{1+(p+j)^2} \quad \in \mathbb{R}[t].$$

Proof. We proceed via induction over $k \in \mathbb{N}_0$.

<u>$k = 0$:</u> We have

$$\partial_t s_{0,p}(t) = \partial_t \int_{\mathbb{R}} u(x,t)^p \,\mathrm{d}x = -p \int_{\mathbb{R}} u(x,t)^p \cdot \partial_x u(x,t)\,\mathrm{d}x$$

with partial integration since $u(\,\cdot\,,t)$ is a Schwartz function

$$\begin{aligned} &= p \int_{\mathbb{R}} \partial_x[u(x,t)^p] \cdot u(x,t)\,\mathrm{d}x = p^2 \int_{\mathbb{R}} u(x,t)^p \cdot \partial_x u(x,t)\,\mathrm{d}x \\ &= -p \cdot \partial_t s_{0,p}(t) \end{aligned}$$

which gives $\partial_t s_{0,p}(t) = 0$ and therefore $s_{0,p}(t) = s_{0,p}(0)$.

<u>$k-1 \to k$:</u> We have

$$\begin{aligned} \partial_t s_{k,p}(t) &= \partial_t \int_{\mathbb{R}} x^k \cdot u(x,t)^p \,\mathrm{d}x \\ &= -p \int_{\mathbb{R}} x^k \cdot u(x,t)^p \cdot \partial_x u(x,t)\,\mathrm{d}x \\ &= p \int_{\mathbb{R}} \partial_x(x^k \cdot u(x,t)^p) \cdot u(x,t)\,\mathrm{d}x \\ &= p\cdot k \int_{\mathbb{R}} x^{k-1} \cdot u(x,t)^{p+1}\,\mathrm{d}x + p^2 \int_{\mathbb{R}} x^k \cdot u(x,t)^p \cdot \partial_x u(x,t)\,\mathrm{d}x \\ &= p\cdot k \cdot s_{k-1,p+1}(t) - p^2 \cdot \partial_t s_{k,p}(t) \\ &= \frac{p\cdot k}{1+p^2} \cdot s_{k-1,p+1}(t) \end{aligned}$$

and solving this induction gives

$$\begin{aligned} s_{k,p}(t) &= s_{k,p}(0) + \frac{p\cdot k}{1+p^2} \int_0^t s_{k-1,p+1}(\tau_1)\,\mathrm{d}\tau_1 \\ &= s_{k,p}(0) + \frac{p\cdot k}{1+p^2} \int_0^t \left[s_{k-1,p+1}(0) + \frac{(p+1)(k-1)}{1+(p+1)^2} \int_0^{\tau_1} s_{k-2,p+2}(\tau_2)\,\mathrm{d}\tau_2 \right] \mathrm{d}\tau_1 \\ &\;\;\vdots \end{aligned}$$

$$= \sum_{i=0}^{k} \frac{s_{k-i,p+i}(0)}{i!} \cdot t^i \cdot \prod_{j=0}^{i-1} \frac{(p+j)\cdot(k-j)}{1+(p+j)^2}$$

which proves the statement. □

Example 12.3.3. For $p = 1$ we have the following three explicit time-dependent moments from Theorem 12.3.2:

$$\begin{aligned} \int_{\mathbb{R}} u(x,t)\ \mathrm{d}x &= s_{0,1}(t) &&= s_{0,1}(0),\\ \int_{\mathbb{R}} x\cdot u(x,t)\ \mathrm{d}x &= s_{1,1}(t) &&= s_{1,1}(0) + s_{0,2}(0)\cdot t,\\ \int_{\mathbb{R}} x^2\cdot u(x,t)\ \mathrm{d}x &= s_{2,1}(t) &&= s_{2,1}(0) + s_{1,2}(0)\cdot t + \frac{2s_{0,3}(0)}{5}\cdot t^2. \end{aligned}$$

For the function

$$u_0(x) := \begin{cases} 1+x & \text{for } x\in[-1,0],\\ 1-x & \text{for } x\in[0,1],\\ 0 & \text{else}\end{cases}$$

we have $s_{0,1}(0) = 1$, $s_{1,1}(0) = 0$, $s_{0,2}(0) = \frac{2}{3}$, $s_{2,1}(0) = \frac{1}{6}$, $s_{1,2}(0) = 0$, $s_{0,3} = \frac{1}{2}$ and therefore

$$\int_{\mathbb{R}} (x-t)^2\cdot u(x,t)\ \mathrm{d}x = L_{s(t)}((x-t)^2) = \frac{1}{6} - \frac{2}{15}t^2 \quad \xrightarrow{t\to\pm\infty} \quad -\infty. \tag{40}$$

Since $u_0 \notin \mathcal{S}(\mathbb{R})$ using a mollifier we get $u_0^\varepsilon := S_\varepsilon * u_0 \in C_0^\infty(\mathbb{R}) \subset \mathcal{S}(\mathbb{R})$ for any $\varepsilon > 0$. We can chose by continuity of the $s_{p,k}(0)$ on ε an $\varepsilon > 0$ small such that the coefficient of t^2 in (40) remains negative. Hence, non-negativity in the assumed classical solution is not preserved, i.e., we have a finite breakdown. ◦

Let $k \in \mathbb{N}$ and $k \geq 2$. For

$$\partial_t u = u^k \cdot \partial_x u \tag{41}$$

multiply (41) with $k\cdot u^{k-1}$ to get $\partial_t(u^k) = u^k\cdot\partial_x(u^k)$. This is Burgers' equation with $v = u^k$. If $u_0 \geq 0$ we can allow $k\in[1,\infty)$ in (41).

13. Transformations of Moment Functionals

The results presented here are published in [dD23b].

Besides the one-point evaluation $L_{x_0}(f) = f(x_0)$ the following is probably the simplest moment functional.

Example 13.0.1. Let λ be the Lebesgue measure on $[0,1]$ and let $\mathcal{V} = \mathbb{R}[t]$. Then the functional

$$L_{\mathrm{Leb}}: \mathbb{R}[t]\to\mathbb{R} \quad \text{with} \quad L_{\mathrm{Leb}}(t^d) = \int_0^1 t^d\ \mathrm{d}\lambda(t) = \frac{1}{d+1} \quad \text{for all } d\in\mathbb{N}_0 \tag{42}$$

is the unique linear functional such that $L(t^d) = \frac{1}{d+1}$ holds for all $d\in\mathbb{N}_0$. ◦

We have seen in Hausdorff's Theorem 3.4.2 that the $[0,1]$-moment problem is one of the simplest to decide. In this chapter we follow the monographs [Fed69], [LL01], and especially [Bog07] for the measure theory and Lebesgue integral. Among other things we want to show that every moment functional has the following form.

Theorem 13.0.2. *Let S be a Souslin set (e.g. a Borel set $S \subseteq \mathbb{R}^n$), $\mathcal{V}$ be a vector space of real measurable functions $v : S \to \mathbb{R}$, and $L : \mathcal{V} \to \mathbb{R}$ be a linear functional. Then the following are equivalent:*

(i) $L : \mathcal{V} \to \mathbb{R}$ is a S-moment functional.

(ii) There exists a measurable function $f : [0,1] \to S$ such that

$$L(v) = \int_0^1 v(f(t)) \, \mathrm{d}\lambda(t) \tag{43}$$

for all $v \in \mathcal{V}$ where λ is the Lebesgue measure on $[0,1]$, i.e., $\lambda \circ f^{-1}$ is a representing measure of L.

13.1. Souslin Sets, Isomorphisms, and Lebesgue–Rohlin Spaces

We have the following transformation formula.

Lemma 13.1.1. *Let $f : (\mathcal{Y}, \mathcal{B}) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ and $g : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ be measurable functions, μ be a measure on $(\mathcal{X}, \mathcal{A})$ such that $f \circ g$ is μ-integrable. Then $\mu \circ g^{-1}$ is a measure on $(\mathcal{Y}, \mathcal{B})$ and f is $\mu \circ g^{-1}$-integrable with*

$$\int_{\mathcal{X}} (f \circ g)(x) \, \mathrm{d}\mu(x) = \int_{\mathcal{Y}} f(y) \, \mathrm{d}(\mu \circ g^{-1})(y). \tag{44}$$

Proof. It is sufficient to show (44) for $f \geq 0$:

$$\begin{aligned}
\int_{\mathcal{X}} (f \circ g)(x) \, \mathrm{d}\mu(x) &= \int_0^\infty \mu((f \circ g)^{-1}((t,\infty))) \, \mathrm{d}t \\
&= \int_0^\infty \mu(g^{-1}(f^{-1}((t,\infty)))) \, \mathrm{d}t \\
&= \int_0^\infty (\mu \circ g^{-1})(f^{-1}((t,\infty))) \, \mathrm{d}t \\
&= \int_{\mathcal{Y}} f(y) \, \mathrm{d}(\mu \circ g^{-1})(y).
\end{aligned}$$

□

Proposition 13.1.2 (see e.g. [Bog07, Prop. 9.1.11])**.** *Let μ be an atomless probability measure on a measurable space $(\mathcal{X}, \mathcal{A})$. Then there exists an $\mathcal{A}$-measurable function $f : \mathcal{X} \to [0,1]$ such that $\mu \circ f^{-1} = \lambda$ is the Lebesgue measure on $[0,1]$.*

Definition 13.1.3 ([Bog07, Def. 6.6.1])**.** A set in a Hausdorff space is called a *Souslin set* if it is the image of a complete separable metric space under a continuous mapping. A *Souslin space* is a Hausdorff space that is a Souslin set.

The empty set is a Souslin set. Souslin sets are fully characterized.

Proposition 13.1.4 (see e.g. [Bog07, Prop. 6.6.3])**.** *Every non-empty Souslin set is the image of* $[0,1] \setminus \mathbb{Q}$ *under some continuous function and also the image of* $(0,1)$ *under some Borel*[29] *mapping.*

More concrete examples which are important to us are the following.

Example 13.1.5. The unit interval $[0,1] \subset \mathbb{R}$ is of course a complete separable metric space (with the usual distance metric $d(x,y) := |x-y|$). The question which sets are the continuous images of $[0,1]$ is partially answered by *space filling curves*, see e.g. [Sag94, Ch. 5]. So the *Peano curves* as continuous and surjective functions

$$f : [0,1] \to [a_1,b_1] \times \cdots \times [a_n,b_n]$$

with $n \in \mathbb{N}$ and $-\infty < a_i < b_i < \infty$ for all $i = 1,\dots,n$ show that all hyper-rectangles are Souslin spaces/sets. Especially $[0,1]$ is a Souslin set/space.

A full answer gives the following theorem.

Hahn–Mazurkiewicz' Theorem 13.1.6 (see [Hah14, Maz20] or e.g. [Sag94, Thm. 6.8])**.** *A set* K *in a non-empty Hausdorff space is the continuous image of* $[0,1]$ *if and only if it is compact, connected, and locally connected.*

So sets $K \subseteq \mathbb{R}^n$ are continuous images of $[0,1]$ if and only if they are compact and path-connected. Hahn–Mazurkiewicz also implies that $\mathbb{PR}^n$ is a Souslin space. ∘

More Souslin sets can be constructed or identified by the following lemma.

Lemma 13.1.7 (see e.g. [Bog07, Lem. 6.6.5, Thm. 6.6.6 and 6.7.3])**.**

(i) The image of a Souslin set under a continuous function to a Hausdorff space is a Souslin set.

(ii) Every open or closed set of a Souslin space is Souslin.

(iii) If A_n *are Souslin sets in* $\mathcal{X}_n$ *for all* $n \in \mathbb{N}$ *then* $\prod_{n\in\mathbb{N}} A_n$ *is a Souslin set in* $\prod_{n\in\mathbb{N}} \mathcal{X}_n$*.*

(iv) If $A_n \subseteq \mathcal{X}$ *are Souslin sets in a Hausdorff space* $\mathcal{X}$*, then* $\bigcap_{n\in\mathbb{N}} A_n$ *and* $\bigcup_{n\in\mathbb{N}} A_n$ *are Souslin sets.*

(v) Every Borel subset of a Souslin space is a Souslin space.

(vi) Let $A \subseteq \mathcal{X}$ *and* $B \subseteq \mathcal{Y}$ *be Souslin sets of Souslin spaces and* $f : \mathcal{X} \to \mathcal{Y}$ *be a Borel function. Then* $f(A)$ *and* $f^{-1}(B)$ *are Souslin sets.*

Remark 13.1.8. The reverse of Lemma 13.1.7(v) is in general not true. Not every Souslin set is Borel. In fact, every non-empty complete metric space without isolated points contains a non-Borel Souslin set, see e.g. [Bog07, Cor. 6.7.11]. ∘

[29] The inverse of open, closed, and Borel sets are Borel sets.

Example 13.1.9. $\mathbb{R}^n$ and every compact semi-algebraic set in $\mathbb{R}^n$ (resp. $\mathbb{PR}^n$) are Souslin sets. ◦

Definition 13.1.10. Let $(\mathcal{X}, \mathcal{A})$ and $(\mathcal{Y}, \mathcal{B})$ be two measurable spaces. A measurable function $\iota : (\mathcal{X}, \mathcal{A}) \to (\mathcal{Y}, \mathcal{B})$ is called an *isomorphism* and the two measurable spaces *isomorphic* if ι is bijective, $\iota(\mathcal{A}) = \mathcal{B}$, and $\iota^{-1}(\mathcal{B}) = \mathcal{A}$.

The reason why we work with Souslin spaces is revealed in the following theorem.

Theorem 13.1.11 (see e.g. [Bog07, Thm. 6.7.4])**.** *Let $\mathcal{X}$ be a Souslin space. Then there exist a Souslin set $S \subseteq [0, 1]$ and an isomorphism $\iota : (S, \mathfrak{B}(S)) \to (\mathcal{X}, \mathfrak{B}(\mathcal{X}))$.*

The existence of an isomorphism can be weakened. For Borel measurable function $f : \mathcal{X} \to \mathcal{Y}$ between two Souslin spaces $\mathcal{X}$ and $\mathcal{Y}$ with $f(\mathcal{X}) = \mathcal{Y}$ one always finds nice (i.e., Borel measurable) one-sided inverse functions.

Jankoff's Theorem 13.1.12 (see [Jan41] or e.g. [Bog07, Thm. 6.9.1 and 9.1.3])**.** *Let $\mathcal{X}$ and $\mathcal{Y}$ be two Souslin spaces and let $f : \mathcal{X} \to \mathcal{Y}$ be a surjective Borel mapping. Then there exists a Borel measurable function $g : \mathcal{Y} \to \mathcal{X}$ such that $f(g(y)) = y$ for all $y \in \mathcal{Y}$.*

In other words, restricting f so some $\mathcal{X}_0 \subseteq \mathcal{X}$ makes $\tilde{f} := f|_{\mathcal{X}_0}$ not only bijective but $\tilde{f}$ and $\tilde{f}^{-1}$ are measurable. We have

$$\mathcal{Y} \xrightarrow{g} \mathcal{X} \xrightarrow{f} \mathcal{Y} \qquad \text{with} \qquad f \circ g = \mathrm{id}_{\mathcal{Y}},$$

i.e., g is injective, f is surjective, and with $\mathcal{X}_0 = \operatorname{im} g := g(\mathcal{Y})$ we have $\tilde{f}^{-1} = g$.

Definition 13.1.13 (see e.g. [Bog07, Def. 9.2.1])**.** Let $(\mathcal{X}, \mathcal{A}, \mu)$ and $(\mathcal{Y}, \mathcal{B}, \nu)$ be two measure spaces with non-negative measures.

i) A *point isomorphism* $T : \mathcal{X} \to \mathcal{Y}$ is a bijective mapping such that $T(\mathcal{A}) = \mathcal{B}$ and $\mu \circ T^{-1} = \nu$.

ii) The spaces $(\mathcal{X}, \mathcal{A}, \mu)$ and $(\mathcal{Y}, \mathcal{B}, \nu)$ are called *isomorphic* mod0 if there exist sets $N \in \mathcal{A}_\mu$, $M \in \mathcal{B}_\nu$ with $\mu(N) = \nu(M) = 0$ and a point isomorphism $T : \mathcal{X} \setminus N \to \mathcal{Y} \setminus M$ that are equipped with the restriction of the measures μ and ν and the σ-algebras $\mathcal{A}_\mu$ and $\mathcal{B}_\nu$.

A point isomorphism T between $(\mathcal{X}, \mathcal{A}, \mu)$ and $(\mathcal{Y}, \mathcal{B}, \nu)$ is of course measurable since $\nu(B) = (\mu \circ T^{-1})(B) = \mu(T^{-1}(B))$ implies $T^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}$.

Like Theorem 13.1.11 also the next result shows the importance of working on Souslin sets.

Theorem 13.1.14 (see e.g. [Bog07, Thm. 9.2.2])**.** *Let $(\mathcal{X}, \mathcal{A})$ be a Souslin space with Borel probability measure μ. Then $(\mathcal{X}, \mathcal{A}, \mu)$ is isomorphic* mod0 *to the space* $([0, 1], \mathfrak{B}([0, 1]), \nu)$ *for some ν Borel probability measure. If μ is an atomless measure, then one can take for ν the Lebesgue measure λ.*

Corollary 13.1.15 (see e.g. [Bog07, Rem. 9.7.4])**.** *Let μ be a probability measure on a Souslin space $\mathcal{X}$. Then there exists a measurable function $f : [0,1] \to \mathcal{X}$ such that $\mu = \lambda \circ f^{-1}$ where λ is the Lebesgue measure on $[0,1]$.*

For both results note the difference to Proposition 13.1.2. In Proposition 13.1.2 we find for any measurable space $\mathcal{X}$ and measure μ a map

$$f : \mathcal{X} \to [0,1] \qquad \text{such that} \qquad \mu = \lambda \circ f^{-1}.$$

But for Souslin spaces $\mathcal{X}$ in Corollary 13.1.15 we find a map

$$f : [0,1] \to \mathcal{X} \qquad \text{such that} \qquad \lambda = \mu \circ f^{-1}.$$

Theorem 13.1.14 restricts $f : [0,1] \to \mathcal{X}$ to isomorphisms and hence not all measures can be transformed into λ. Atoms in the measure μ prevent it from being isomorphic to λ. In fact, as explained in [Bog07, Rem. 9.7.4], Corollary 13.1.15 follows from Theorem 13.1.14 by introducing atoms into $f : [0,1] \to \mathcal{X}$ by introducing constant functions into f.

But Theorem 13.1.14 provides that if μ has atoms, it can still be isomorphic mod0 transformed into a measure ν on $[0,1]$. Without atoms we could chose $\nu = \lambda$. So is it possible to transform the non-atomic part of μ to λ and then add the atoms from μ to λ? Yes, we can. This is done on the following spaces.

Definition 13.1.16 (see e.g. [Bog07, Def. 9.4.6])**.** A measure space $(\mathcal{X}, \mathcal{A}, \mu)$ is called a *Lebesgue–Rohlin space* if it is isomorphic mod0 to some measure space $(\mathcal{Y}, \mathcal{B}, \nu)$ with a countable basis with respect to which $\mathcal{Y}$ is complete.

Example 13.1.17 (see e.g. [Bog07, Exm. 9.4.2])**.** $(M, \mathfrak{B}(M), \mu)$, where M is a Borel set of a complete separable metric space $\mathcal{X}$ and μ is a Borel measure on M, is a Lebesgue–Rohlin space. Especially $\mathcal{X} = \mathbb{R}^n$ or $\mathbb{PR}^n$ are complete metric spaces and therefore any Borel measure on a Borel subset $M \in \mathfrak{B}(\mathbb{R}^n)$ gives a Lebesgue–Rohlin space. ∘

We can now transform any measure by an isomorphism mod0 to the Lebesgue measure λ plus atoms.

Theorem 13.1.18 (see e.g. [Bog07, Thm. 9.4.7])**.** *Let $(\mathcal{X}, \mathcal{A}, \mu)$ be a Lebesgue–Rohlin space with a probability measure μ. Then it is isomorphic* mod0 *to the interval* $[0,1]$ *with the measure $\nu = c\lambda + \sum_{i=1}^{\infty} c_n \cdot \delta_{1/n}$, where $c = 1 - \sum_{i=1}^{\infty} c_i$, $\mu(a_i) = c_i$ and $\{a_i\} \subseteq \mathcal{X}$ is the family of all atoms of μ.*

So we can transform any measure to the Lebesgue measure λ on $[0,1]$ or to λ on $[0,1]$ plus atoms. But these transformations are performed mainly by measurable functions because the set $\mathcal{X}$ where the original measure lives is too large. If we restrict the space where the measure lives, we get better transformations, especially continuous ones.

Kolesnikov's Theorem 13.1.19 (see [Kol99] or e.g. [Bog07, Thm. 9.7.1])**.** *Let K be a compact metric space that is the image of* $[0,1]$ *under a continuous mapping $\tilde{f}$ and let μ be a Borel probability measure on K such that* $\operatorname{supp}\mu = K$. *Then there exists a continuous and surjective mapping $f : [0,1] \to K$ such that $\mu = \lambda \circ f^{-1}$, λ is the Lebesgue measure on* $[0,1]$.

We will apply Kolesnikov's Theorem 13.1.19 especially in connection with the Hahn–Mazurkiewicz' Theorem 13.1.6. The advantage is here that f on $[0,1]$ is continuous and can therefore be approximated by polynomials up to any precision $\varepsilon > 0$ in the sup-norm.

13.2. Transformations of Moment Functionals

Proof of Theorem 13.0.2. (i)→(ii): Let μ be a representing measure of L. By Corollary 13.1.15 there exists a measurable function $f : [0,1] \to S$ such that $\mu = \lambda \circ f^{-1}$ and hence

$$L(v) = \int_S v(x)\ \mathrm{d}\mu(x) = \int_S v(x)\ \mathrm{d}(\lambda \circ f^{-1})(x) \overset{\text{Lemma 13.1.1}}{=} \int_0^1 v(f(t))\ \mathrm{d}\lambda(t)$$

for all $v \in \mathcal{V}$.

(ii)→(i): $\lambda \circ f^{-1}$ is a representing measure of L by Lemma 13.1.1. □

Definition 13.2.1. Let $\mathcal{X}$ and $\mathcal{Y}$ be two Souslin spaces, $\mathcal{U}$ and $\mathcal{V}$ two vector spaces of real measurable functions on $\mathcal{X}$ resp. $\mathcal{Y}$, and $K : \mathcal{U} \to \mathbb{R}$ and $L : \mathcal{V} \to \mathbb{R}$ be two linear functionals. We say L *(continuously) transforms into* K, symbolized by $L \rightsquigarrow K$ resp. $L \overset{c}{\rightsquigarrow} K$, if there exists a Borel (resp. continuous) function $f : \mathcal{X} \to \mathcal{Y}$ such that $\mathcal{V} \circ f \subseteq \mathcal{U}$ and $L(v) = K(v \circ f)$ for all $v \in \mathcal{V}$.

We say L *strongly (and continuously) transforms into* K, symbolized by $L \overset{s}{\rightsquigarrow} K$ resp. $L \overset{sc}{\rightsquigarrow} K$, if there exists a surjective Borel (resp. surjective and continuous) function $f : \mathcal{X} \twoheadrightarrow \mathcal{Y}$ such that $\mathcal{V} \circ f = \mathcal{U}$ and $L(v) = K(v \circ f)$ for all $v \in \mathcal{V}$.

Corollary 13.2.2. *$L : \mathcal{V} \to \mathbb{R}$ is a moment functional iff $L \rightsquigarrow [K : \mathcal{L}^1([0,1], \lambda) \to \mathbb{R}]$.*

For the transformation $\rightsquigarrow$ between two linear functionals in Definition 13.2.1 we get the following technical result.

Lemma 13.2.3. *Let $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ be Souslin spaces; $\mathcal{U}$, $\mathcal{V}$, and $\mathcal{W}$ be vector spaces of real measurable functions on $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ respectively; and $M : \mathcal{W} \to \mathbb{R}$, $L : \mathcal{V} \to \mathbb{R}$, and $K : \mathcal{U} \to \mathbb{R}$ be linear functionals. The following hold:*

(i) $M \rightsquigarrow L$ and $L \rightsquigarrow K$ imply $M \rightsquigarrow K$.

(ii) $M \overset{c}{\rightsquigarrow} L$ and $L \overset{c}{\rightsquigarrow} K$ imply $M \overset{c}{\rightsquigarrow} K$.

(iii) $M \overset{s}{\rightsquigarrow} L$ and $L \overset{s}{\rightsquigarrow} K$ imply $M \overset{s}{\rightsquigarrow} K$.

(iv) $M \overset{sc}{\rightsquigarrow} L$ and $L \overset{sc}{\rightsquigarrow} K$ imply $M \overset{sc}{\rightsquigarrow} K$.

Proof. (i): Since $M \rightsquigarrow L$ there exists a Borel function $f : \mathcal{Y} \to \mathcal{Z}$ such that $\mathcal{W} \circ f \subseteq \mathcal{V}$ and $M(w) = L(w \circ f)$ for all $w \in \mathcal{W}$. And since $L \rightsquigarrow K$ there exists a Borel function $g : \mathcal{X} \to \mathcal{Y}$ such that $\mathcal{V} \circ g \subseteq \mathcal{U}$ and $L(v) = K(v \circ g)$ for all $v \in \mathcal{V}$. Hence, $h = f \circ g : \mathcal{X} \to \mathcal{Z}$ implies $\mathcal{W} \circ h = \mathcal{W} \circ f \circ g \subseteq \mathcal{V} \circ g \subseteq \mathcal{U}$ and $M(w) = L(w \circ f) = K(w \circ f \circ g) = K(w \circ h)$ for all $w \in \mathcal{W}$, i.e., $M \rightsquigarrow K$.

(ii)-(iv) follow in the same way as (i). □

Lemma 13.2.3 can be seen as shortening the sequence:

$$M \rightsquigarrow L \rightsquigarrow K \qquad \Rightarrow \qquad M \rightsquigarrow K.$$

Lemma 13.2.4. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Souslin sets, $\mathcal{U}$ and $\mathcal{V}$ vector spaces of real functions on $\mathcal{X}$ resp. $\mathcal{Y}$, and $L : \mathcal{V} \to \mathbb{R}$ and $K : \mathcal{U} \to \mathbb{R}$ be linear functionals. Then $L \overset{s}{\rightsquigarrow} K$ implies $K \rightsquigarrow L$.*

Proof. Since $L \overset{s}{\rightsquigarrow} K$ there exists a surjective Borel function $f : \mathcal{X} \to \mathcal{Y}$ such that $L(v) = K(v \circ f)$ and $\mathcal{V} \circ f = \mathcal{U}$. Since f is surjective by Jankoff's Theorem 13.1.12 there exists a Borel function $g : \mathcal{Y} \to \mathcal{X}$ such that $f(g(y)) = y$ for all $y \in \mathcal{Y}$. Let $u \in \mathcal{U} = \mathcal{V} \circ f$, then v in $u = v \circ f$ is unique since for v_1 and v_2 with that property we have

$$v_1 = v_1 \circ f \circ g = u \circ g = v_2 \circ f \circ g = v_2.$$

Hence, $\mathcal{U} \circ g = \mathcal{V}$ and for all $u \in \mathcal{U}$ we have

$$K(u) = K(v \circ f) = L(v) = L(v \circ f \circ g) = L(u \circ g). \qquad \square$$

Theorem 13.2.5. *Let $\mathcal{X}$ and $\mathcal{Y}$ be Souslin sets, $\mathcal{U}$ and $\mathcal{V}$ vector spaces of real functions on $\mathcal{X}$ resp. $\mathcal{Y}$, and $L\colon \mathcal{V} \to \mathbb{R}$ and $K\colon \mathcal{U} \to \mathbb{R}$ be linear functionals. If $L \rightsquigarrow K$, then*

(i) K is a moment functional

implies

(ii) L is a moment functional.

If $L \overset{s}{\rightsquigarrow} K$, then (i) $\Leftrightarrow$ (ii).

Proof. Since $L \rightsquigarrow K$ there exists a Borel function $f : \mathcal{X} \to \mathcal{Y}$ such that $\mathcal{V} \circ f \subseteq \mathcal{U}$ and $L(v) = K(v \circ f)$ for all $v \in \mathcal{V}$.

(i)$\to$(ii): Let K be a moment functional with representing measure ν on $\mathcal{X}$, then

$$L(v) = K(v \circ f) = \int_{\mathcal{X}} (v \circ f)(x) \; \mathrm{d}\nu(x) \overset{\text{Lemma 13.1.1}}{=} \int_{\mathcal{Y}} v(y) \; \mathrm{d}(\nu \circ f^{-1})(y),$$

i.e., $\nu \circ f^{-1}$ is a representing measure of L and hence L is a moment functional.

(ii)$\to$(i): When $L \overset{s}{\rightsquigarrow} K$, then Lemma 13.2.4 implies $K \rightsquigarrow L$. $\square$

The importance of the transformation and hence Theorem 13.2.5 can be seen in

$$\begin{tikzcd}
 & L_8 \arrow[d, rightsquigarrow] & L_6 \arrow[r, rightsquigarrow] & L_5 \arrow[d, rightsquigarrow] & \\
L_4 \arrow[r, rightsquigarrow] & L_3 \arrow[r, rightsquigarrow] & L_2 \arrow[r, rightsquigarrow] & L_1 \arrow[r, rightsquigarrow] & K. \\
 & & L_7 \arrow[u, rightsquigarrow] & &
\end{tikzcd} \tag{45}$$

If K is a moment functional, then all $L_1, \dots, L_8$ are moment funtionals. Assume in (45) all transformations $\rightsquigarrow$ are strong transformations $\overset{s}{\rightsquigarrow}$. Then: If one L_i or K is a moment functional, then all $K, L_1, \dots, L_8$ are moment functionals.

Proposition 13.2.6. *Let $\mathcal{V}$ be a vector space of real measurable functions on a measurable space $(\mathcal{X},\mathcal{A})$ such that there exists an element $v\in\mathcal{V}$ with $1\leq v$ on $\mathcal{X}$ and let $L:\mathcal{V}\to\mathbb{R}$ be a moment functional which has an atomless representing measure. Then there exists a measurable function $f:\mathcal{X}\to[0,1]$ and an extension $\overline{L}:\mathcal{V}+\mathbb{R}[f]\to\mathbb{R}$ of L such that $\overline{L}(f^d)=\frac{\overline{L}(1)}{d+1}$ for all $d\in\mathbb{N}_0$, i.e., $\tilde{L}:\mathbb{R}[t]\to\mathbb{R}$ with $\tilde{L}(t^d):=\overline{L}(f^d)$ for all $d\in\mathbb{N}_0$ is represented by $\overline{L}(1)\cdot\lambda$ where λ is the Lebesgue measure λ on $[0,1]$.*

Proof. Let μ be a representing measure of L. By Proposition 13.1.2 there exists a measurable $f:\mathbb{R}^n\to[0,1]$ such that $\mu\circ f^{-1}=\lambda$ on $[0,1]$. Since f is measurable, $|f|\leq 1$ on $\mathbb{R}^n$, and $L(1)<\infty$, all f^d, $d\in\mathbb{N}_0$, are μ-integrable:

$$\left|\int_{\mathbb{R}^n} f^d(x)\ \mathrm{d}\mu(x)\right|\leq\int_{\mathbb{R}^n}|f(x)|^d\ \mathrm{d}\mu(x)\leq\int_{\mathbb{R}^n}1\ \mathrm{d}\mu(x)=L(1).$$

Define $\overline{L}:\mathbb{R}[f]\to\mathbb{R}$ by $\overline{L}(f^d):=\int_{\mathbb{R}^n}f^d(x)\ \mathrm{d}\mu(x)$. Then

$$\overline{L}(f^d)=\int_{\mathbb{R}^n}f^d(x)\ \mathrm{d}\mu(x)\overset{\text{Lemma 13.1.1}}{=}\int_0^1 t^d\ \mathrm{d}(\mu\circ f^{-1})(t)=\int_0^1 t^d\ \mathrm{d}\lambda(t)=\frac{L(1)}{d+1}$$

is represented by $L(1)\cdot\lambda$ on $[0,1]$. □

Theorem 13.2.7. *Let $n\in\mathbb{N}$ be a natural number, $K\subset\mathbb{R}^n$ be a compact and path-connected set, and let $\mathcal{V}$ be a vector space of real measurable functions on $(K,\mathfrak{B}(K))$. Then any surjective and continuous function $f:[0,1]\to K$ induces for any linear functional $L:\mathcal{V}\to\mathbb{R}$ a strong and continuous transformation*

$$L:\mathcal{V}\to\mathbb{R}\quad\overset{sc:f}{\rightsquigarrow}\quad\tilde{L}:\mathcal{V}\circ f\to\mathbb{R},$$

i.e., for any linear functional $L:\mathcal{V}\to\mathbb{R}$ the following are equivalent:

(i) $L:\mathcal{V}\to\mathbb{R}$ is a K-moment functional.

(ii) $\tilde{L}:\mathcal{V}\circ f\to\mathbb{R}$ defined by $\tilde{L}(v\circ f):=L(v)$ is a $[0,1]$-moment functional.

If $\tilde{\mu}$ is a representing measure of $\tilde{L}$, then $\tilde{\mu}\circ f^{-1}$ is a representing measure of L.

There exists a measurable function $g:K\to[0,1]$ such that $f(g(x))=x$ for all $x\in K$ and if μ is a representing measure of L, then $\mu\circ g^{-1}$ is a representing measure of $\tilde{L}$.

Proof. Since $K\subset\mathbb{R}^n$ is compact and path-connected, by the Hahn–Mazurkiewicz' Theorem 13.1.6 there exists a continuous and surjective function $f:[0,1]\to K$. By Example 13.1.5 or Lemma 13.1.7 $[0,1]$ and K are Souslin spaces and f is Borel measurable (since it is continuous). By Jankoff's Theorem 13.1.12 there exists a measurable function $g:K\to[0,1]$ such that

$$f(g(x))=x\quad\text{for all } x\in K.\tag{46}$$

(46) implies that $\tilde{L}$ is well-defined by $\tilde{L}(v\circ f)=L(v)$. To show this, for $\tilde{v}\in\tilde{\mathcal{V}}$ let $v_1,v_2\in\mathcal{V}$ be such that $v_1\circ f=\tilde{v}=v_2\circ f$. But then g resp. (46) implies $v_1=v_1\circ f\circ g=\tilde{v}\circ g=v_2\circ f\circ g=v_2$, i.e., for any $\tilde{v}\in\mathcal{V}$ there is a unique $v\in\mathcal{V}$ with $\tilde{v}=v\circ f$.

(i)→(ii): Let $L : \mathcal{V} \to \mathbb{R}$ be a K-moment functional and μ be a representing measure of L, i.e., $\operatorname{supp}\mu \subseteq K$ and

$$L(v) = \int_K v(x)\ \mathrm{d}\mu(x) \quad \text{for all } v \in \mathcal{V}.$$

Then

$$\begin{aligned}\tilde{L}(v \circ f) = L(v) = \int_K v(x)\ \mathrm{d}\mu(x) \quad &= \quad \int_K (v \circ f)(g(x))\ \mathrm{d}\mu(x) \\ &\overset{\text{Lemma 13.1.1}}{=} \int_0^1 (v \circ f)(y)\ \mathrm{d}(\mu \circ g^{-1})(y),\end{aligned}$$

i.e., $\mu \circ g^{-1}$ is a representing measure of $\tilde{L}$ and hence $\tilde{L}$ is a $[0,1]$-moment functional.

(ii)→(i): Let $\tilde{\mu}$ be a representing measure of $\tilde{L} : \tilde{\mathcal{V}} \to \mathbb{R}$. Then

$$L(v) = \tilde{L}(v \circ f) = \int_0^1 (v \circ f)(y)\ \mathrm{d}\tilde{\mu}(y) \overset{\text{Lemma 13.1.1}}{=} \int_K v(x)\ \mathrm{d}(\tilde{\mu} \circ f^{-1})(x),$$

i.e., $\tilde{\mu} \circ f^{-1}$ is a representing measure of L with $\operatorname{supp}\tilde{\mu} \circ f^{-1} \subseteq K$ and L is therefore a K-moment sequence. □

Corollary 13.2.8. *Let $n \in \mathbb{N}$ and $K \subset \mathbb{R}^n$ be the union of $k \in \mathbb{N} \cup \{\infty\}$ compact, path-connected and pairwise disjoint sets $K_i \subset \mathbb{R}^n$: $K = \bigcup_{i=1}^k K_i$. Let $\mathcal{V}$ be a vector space of real valued measurable functions on $(K, \mathfrak{B}(K))$. There exists a continuous surjective function*

$$f : \bigcup_{i=1}^k [2i-2, 2i-1] \to K$$

such that for any linear functional $L : \mathcal{V} \to \mathbb{R}$ the following are equivalent:

(i) $L : \mathcal{V} \to \mathbb{R}$ *is a K-moment functional.*

(ii) $\tilde{L} : \tilde{\mathcal{V}} \to \mathbb{R}$ *on $\tilde{\mathcal{V}} := \{v \circ f \mid v \in \mathcal{V}\}$ and defined by $\tilde{L}(v \circ f) := L(v)$ is a $\bigcup_{i=1}^k [2i-2, 2i-1]$-moment functional.*

An advantage in Theorem 13.2.7 is that $f = (f_1, \dots, f_n) : [0,1] \to K \subset \mathbb{R}^n$ is continuous. Hence, all coordinate functions $f_i : [0,1] \to \mathbb{R}$ are continuous. By the Stone–Weierstrass Theorem we can approximate each f_i in the sup-norm on $[0,1]$ by polynomials to any precision. f can therefore be approximated to any precision by a polynomial map. A representing measure $\tilde{\mu}$ of $\tilde{L}$ provides the representing measure $\tilde{\mu} \circ f^{-1}$ of L. An approximation $f_\varepsilon \in \mathbb{R}[x_1, \dots, x_n]^n$ of f, i.e., $\sup_{t \in [0,1]} \|f(t) - f_\varepsilon(t)\| < \varepsilon$ with any (fixed) norm $\|\cdot\|$ on $\mathbb{R}^n$ and $\varepsilon > 0$, provides an approximate representing measure $\tilde{\mu} \circ f_\varepsilon^{-1}$ of L.

Let $K \subset \mathbb{R}^n$ be a compact and path-connected set, $\mathcal{V} = \mathbb{R}[x_1, \dots, x_n]$, and $L : \mathcal{V} \to \mathbb{R}$ be a linear functional. Then the induced functional $\tilde{L} : \tilde{\mathcal{V}} \to \mathbb{R}$ on $[0,1]$ is defined by $\tilde{L}(p \circ f) := L(p)$. It depends on $p \circ f$, i.e., $f^\alpha = f_1^{\alpha_1} \cdots f_n^{\alpha_n}$, $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$. So as in Theorem 13.0.2 the algebraic structure of $\mathbb{R}[x_1,\dots,x_n]$ remains but the domain K is pulled back to $[0,1]$ by the continuous f.

That the algebraic structure remains also reveals one big difference between L and $\tilde{L}$. E.g. $\mathcal{V} = \mathbb{R}[x_1,\dots,x_n]$ separates points and is therefore dense in $C(K,\mathbb{R})$. But $f:[0,1]\to K$ is a space filling curve and therefore never injective (Netto's Theorem). Hence, there are $t_1,t_2\in[0,1]$ with $t_1\neq t_2$ and $f(t_1)=f(t_2)$. The set $\tilde{\mathcal{V}} := \{p\circ f \mid p\in\mathcal{V}\}$ therefore does not separate t_1 from t_2 and is by the Stone–Weierstrass Theorem not dense in $C([0,1],\mathbb{R})$. So the $\tilde{L}$ in Theorem 13.2.7 and Corollary 13.2.8 can at this point not extended to the Hausdorff Moment Problem (Hausdorff's Theorem 3.4.2).

Theorem 13.2.9. *Let $n\in\mathbb{N}$ be a natural number and $K\subset\mathbb{R}^n$ be a compact and path-connected set. Then there exists a measurable function*

$$g : K \to [0,1]$$

such that for all linear functionals $L:\mathcal{V}\to\mathbb{R}$ with $1\in\mathcal{V}\subseteq C(K,\mathbb{R})$ the following are equivalent:

(i) $L:\mathcal{V}\to\mathbb{R}$ is a K-moment functional.

(ii) $L:\mathcal{V}\to\mathbb{R}$ continuously[30] extends to $\overline{L}:\mathcal{V}+\mathbb{R}[g]\to\mathbb{R}$ such that $\tilde{L}:\mathbb{R}[t]\to\mathbb{R}$ defined by $\tilde{L}(t^d) := \overline{L}(g^d)$ for all $d\in\mathbb{N}_0$ is a $[0,1]$-moment functional, i.e.,

$$\begin{array}{ccc} & & L:\mathcal{V}\to\mathbb{R} \\ & & \big\downarrow \mathrm{id}_{\mathcal{X}} \\ \tilde{L}:\mathbb{R}[t]\to\mathbb{R} & \overset{g}{\rightsquigarrow} & \overline{L}:\mathcal{V}+\mathbb{R}[g]\to\mathbb{R}. \end{array} \tag{47}$$

If μ is the representing measure of L, then $\mu\circ g^{-1}$ represents $\tilde{L}$.

Additionally, there exists a continuous and surjective function $f:[0,1]\to K$ independent on L resp. $\tilde{L}$ such that $f(g(x))=x$ for all $x\in K$ and if $\tilde{\mu}$ is the representing measure of $\tilde{L}$, then $\tilde{\mu}\circ f^{-1}$ is the representing measure of L.

Proof. Since K is a compact and path-connected set, by the Hahn–Mazurkiewicz' Theorem 13.1.6 there exists a continuous and surjective function $f:[0,1]\to K$. By Lemma 13.1.7 $[0,1]$ and K are Souslin sets and hence by Jankoff's Theorem 13.1.12 there exists a measurable function $g:K\to[0,1]$ such that

$$f(g(x)) = x \quad \text{for all} \quad x\in K. \tag{48}$$

(i)→(ii): Let $L:\mathcal{V}\to\mathbb{R}$ be a K-moment functional and μ be a representing measure of L with $\operatorname{supp}\mu\subseteq K$. g is measurable with $|g|\leq 1$ and hence we have that all g^d, $d\in\mathbb{N}_0$, are μ-integrable by

$$\left|\int_K g(x)^d\,\mathrm{d}\mu(x)\right| \leq \int_K |g(x)|^d\,\mathrm{d}\mu(x) \leq \int_K 1\,\mathrm{d}\mu(x) = \mu(K) = L(1) \tag{49}$$

[30] If $p_i\in\mathbb{R}[t]$ with $p_i \rightrightarrows p\in C([0,1],\mathbb{R})$ and $p\circ g\in\mathcal{V}$ then $\overline{L}(p_i\circ g)\to L(p\circ g)$.

and hence L extents to $\mathbb{R}[g]$. Let $p \in \mathbb{R}[t]$, then

$$\tilde{L}(p) = L(p \circ g) = \int_K (p \circ g)(x)\ \mathrm{d}\mu(x) \overset{\text{Lemma 13.1.1}}{=} \int_0^1 p(t)\ \mathrm{d}(\mu \circ g^{-1})(t)$$

and $\mu \circ g^{-1}$ is a representing measure of $\tilde{L}$, i.e., $\tilde{L}$ is a $[0,1]$-moment functional.

(ii)→(i): Let $\tilde{L} : \mathbb{R}[t] \to \mathbb{R}$ be a $[0,1]$-moment functional and $\tilde{\mu}$ be its unique representing measure. Since by the Stone–Weierstrass Theorem $\mathbb{R}[t]$ is dense in $C([0,1],\mathbb{R})$ the moment functional $\tilde{L}$ extends uniquely to $C([0,1],\mathbb{R})$. For simplicity we denote this extension also $\tilde{L} : C([0,1],\mathbb{R}) \to \mathbb{R}$. Since $f : [0,1] \to K$ is continuous we have $v \circ f \in C([0,1],\mathbb{R})$ for all $v \in \mathcal{V}$. By (48) we have $v = v \circ f \circ g$ for all $v \in \mathcal{V}$ and hence

$$L(v) = L(v \circ f \circ g). \tag{50}$$

But since $v \circ f : [0,1] \to \mathbb{R}$ is continuous and $\tilde{L} : \mathbb{R}[t] \to \mathbb{R}$ uniquely extends to $C([0,1],\mathbb{R})$ we have

$$L(v \circ f \circ g) = \tilde{L}(v \circ f). \tag{51}$$

In summary we get

$$\begin{aligned} L(v) \overset{(50)}{=} L(v \circ f \circ g) \overset{(51)}{=} \tilde{L}(v \circ f) &= \int_0^1 (v \circ f)(t)\ \mathrm{d}\tilde{\mu}(t) \\ &\overset{\text{Lem. 13.1.1}}{=} \int_K v(x)\ \mathrm{d}(\tilde{\mu} \circ f^{-1})(x) \end{aligned} \tag{52}$$

for all $v \in \mathcal{V}$, i.e., $\tilde{\mu} \circ f^{-1}$ is a representing measure of L and L is therefore a K-moment functional. □

Theorem 13.2.10. *Let $n \in \mathbb{N}$ be a natural number, $K \subset \mathbb{R}^n$ be a compact and path-connected set, and let $g : K \to [0,1]$ be from Theorem 13.2.9. Then for any $\varepsilon > 0$ and K-moment functional $L : \mathbb{R}[x_1,\dots,x_n] \to \mathbb{R}$ there exists a polynomial $g_\varepsilon \in \mathbb{R}[x_1,\dots,x_n]$ such that*

$$L(|g_\varepsilon - g|) \leq \varepsilon \qquad \textit{and} \qquad |L(g^d) - L(g_\varepsilon^d)| \leq d \cdot L(|g - g_\varepsilon|) \leq d \cdot \varepsilon$$

hold for all $d \in \mathbb{N}_0$. g_ε can be chosen to be a square: $g_\varepsilon = p_\varepsilon^2$ for some $p_\varepsilon \in \mathbb{R}[x_1,\dots,x_n]$.

Proof. L is a K-moment functional and therefore has a unique representing measure μ with $\operatorname{supp}\mu \subseteq K$. $g \geq 0$ and hence there exists a measurable function $p : K \to [0,1]$ such that $g = p^2$. Since K is compact and $\mu(K) = L(1) < \infty$ the polynomials $\mathbb{R}[x_1,\dots,x_n]$ are dense in $L^1(K,\mu)$. By

$$\left|\int_K p(x)\ \mathrm{d}\mu(x)\right| \leq \int_K |p(x)|\ \mathrm{d}\mu(x) \leq \int_K 1\ \mathrm{d}\mu(x) = L(1) < \infty$$

we have $p \in L^1(K,\mu)$ and therefore for any $\varepsilon > 0$ there exists a $p_\varepsilon \in \mathbb{R}[x_1,\dots,x_n]$ such that $p_\varepsilon \leq 1$ on K and

$$\|p - p_\varepsilon\|_{L^1(K,\mu)} = \int_K |p(x) - p_\varepsilon(x)|\ \mathrm{d}\mu(x) \leq \frac{1}{2}\varepsilon.$$

Set $g_\varepsilon := p_\varepsilon^2$. Then

$$\begin{aligned} L(|g-g_\varepsilon|) &= \int_K |g-g_\varepsilon| \,\mathrm{d}\mu(x) = \int_K |p^2(x)-p_\varepsilon^2(x)| \,\mathrm{d}\mu(x) \\ &= \int_K |p-p_\varepsilon|\cdot|p+p_\varepsilon| \,\mathrm{d}\mu(x) \leq 2\int_K |p(x)-p_\varepsilon(x)| \,\mathrm{d}\mu(x) \leq \varepsilon. \end{aligned} \tag{53}$$

For $d=0$ we have $g^0 = g_\varepsilon^0 = 1$, i.e., $L(g^0) = L(1) = L(g_\varepsilon^0)$, and for $d=1$ we have $|L(g)-L(g_\varepsilon)| \leq L(|g-g_\varepsilon|) \leq \varepsilon$. So let $d \geq 2$. Then

$$\begin{aligned} |L(g^d)-L(g_\varepsilon^d)| \leq L(|g^d-g_\varepsilon^d|) &= \int_K |g(x)^d - g_\varepsilon(x)^d| \,\mathrm{d}\mu(x) \\ &= \int_K |g(x)-g_\varepsilon(x)|\cdot\left|\sum_{i=0}^{d-1} g(x)^i\cdot g_\varepsilon(x)^{d-1-i}\right| \,\mathrm{d}\mu(x) \qquad (54) \\ &\leq d\cdot\int_K |g(x)-g_\varepsilon(x)| \,\mathrm{d}\mu(x) \leq d\cdot\varepsilon. \end{aligned}$$

□

Corollary 13.2.11. *Let $k,\ n \in \mathbb{N}$ be natural numbers and $K \subset \mathbb{R}^n$ be the union of finitely many compact, path-connected, and pairwise disjoint sets K_i: $K = \bigcup_{i=1}^k K_i$. Then there exists a measurable function*

$$g : K \to I_k := \bigcup_{i=1}^k \left[\frac{2i-2}{2k-1}, \frac{2i-1}{2k-1}\right] \subset [0,1]$$

such that for all linear functionals $L : \mathcal{V} \to \mathbb{R}$ with $1 \in \mathcal{V} \subseteq C(K,\mathbb{R})$ the following are equivalent:

(i) $L : \mathbb{R}[x_1,\dots,x_n] \to \mathbb{R}$ *is a K-moment functional.*

(ii) $L : \mathcal{V} \to \mathbb{R}$ *continuously extends to $\overline{L} : \mathcal{V} + \mathbb{R}[g] \to \mathbb{R}$ such that $\tilde{L} : \mathbb{R}[t] \to \mathbb{R}$ defined by $\tilde{L}(t^d) := \overline{L}(g^d)$ for all $d \in \mathbb{N}_0$ is a $[0,1]$-moment functional.*

Corollary 13.2.12. *Let $n, k \in \mathbb{N}$ be natural numbers, $K \subset \mathbb{R}^n$ the union of finitely many compact, path-connected, and pairwise disjoint sets K_i, $K = \bigcup_{i=1}^k K_i$, and let $g : K \to I_k$ be from Corollary 13.2.11. Then for any $\varepsilon > 0$ and K-moment functional $L : \mathbb{R}[x_1,\dots,x_n] \to \mathbb{R}$ there exists a polynomial $g_\varepsilon \in \mathbb{R}[x_1,\dots,x_n]$ such that*

$$L(|g_\varepsilon - g|) \leq \varepsilon \qquad \text{and} \qquad |L(g^d) - L(g_\varepsilon^d)| \leq d\cdot L(|g-g_\varepsilon|) \leq d\cdot\varepsilon$$

hold for all $d \in \mathbb{N}_0$. g_ε can be chosen to be a square: $g_\varepsilon = p_\varepsilon^2$ for some $p_\varepsilon \in \mathbb{R}[x_1,\dots,x_n]$.

Theorem 13.2.13. *Let $n \in \mathbb{N}$ be a natural number, $B \in \mathfrak{B}(\mathbb{R}^n)$ be a Borel set, and $\mathcal{V}$ be a vector space of real measurable functions on B with $1 \in \mathcal{V}$. Then the following are equivalent.*

(i) $L : \mathcal{V} \to \mathbb{R}$ *is a B-moment functional.*

(ii) There exist Borel sets $M \in \mathfrak{B}(B)$ and $N \in \mathfrak{B}([0,1])$ and a bijective and measurable function (isomorphism) $f : [0,1] \setminus N \to B \setminus M$ such that

$$L(v) = \int_0^1 v(f(t))\ \mathrm{d}\nu(t) \qquad \textit{with} \qquad \nu = c \cdot \lambda + \sum_{i\in\mathbb{N}} c_i \cdot \delta_{1/i} \tag{55}$$

for all $v \in \mathcal{V}$, where c, $c_i \geq 0$ and $c + \sum_{i\in\mathbb{N}} c_i = L(1)$, i.e., $\nu \circ f^{-1}$ is a representing measure of L.

Proof. (ii)→(i): Clear since $\nu \circ f^{-1}$ is a representing measure of L.

(i)→(ii): Let μ be a representing measure of L. Then $(B, \mathfrak{B}(B), \mu)$ is by Example 13.1.17 a Lebesgue–Rohlin space and therefore by Theorem 13.1.18 isomorph mod0 to $([0,1], \mathfrak{B}([0,1]), \nu)$ with ν as in (55), i.e., there exist Borel sets $M \in \mathfrak{B}(B)$ and $N \in \mathfrak{B}([0,1])$ and a bijective and measurable function $f : [0,1]\setminus N \to B\setminus M$ such that $\nu = \mu\circ f$ and $\mu(M) = \nu(N) = 0$. Then by Lemma 13.1.1 for all $v \in \mathcal{V}$ we have

$$\begin{aligned} L(v) &= \int_B v(x)\ \mathrm{d}\mu(x) = \int_{B\setminus M} v(f \circ f^{-1})\ \mathrm{d}\mu(x) \\ &= \int_{[0,1]\setminus N} v(f(t))\ \mathrm{d}(\mu \circ f)(t) = \int_0^1 v(f(t))\ \mathrm{d}\nu(t). \end{aligned}$$

□

Theorem 13.2.14. *Let $n \in \mathbb{N}$, $K \subset \mathbb{R}^n$ be a compact and path-connected set, $\mathcal{V}$ be a vector space of real function on K, and $L : \mathcal{V} \to \mathbb{R}$ be a linear functional. Then the following are equivalent:*

(i) $L : \mathcal{V} \to \mathbb{R}$ is a K-moment functional with representing measure μ such that $\operatorname{supp}\mu = K$.

(ii) There exists a continuous and surjective function $f : [0,1] \to K$ such that

$$L(v) = \int_0^1 v(f(t))\ \mathrm{d}\lambda(t)$$

for all $v \in \mathcal{V}$ where λ is the Lebesgue measure on $[0,1]$, i.e.,

$$L \quad \overset{f}{\rightsquigarrow} \quad L_{Leb} : \mathcal{L}^1([0,1],\lambda) \to \mathbb{R}.$$

Proof. (i)→(ii): Let $L : \mathcal{V} \to \mathbb{R}$ be a K-moment functional and let μ be its unique representing measure with $\operatorname{supp}\mu = K$. Since K is a compact and path-connected set, by the Hahn–Mazurkiewicz' Theorem 13.1.6 there exists a continuous and surjective function $\tilde{f} : [0,1] \to K$. By Kolesnikov's Theorem 13.1.19 there exists a continuous and surjective function $f : [0,1] \to K$ such that $\mu = \lambda \circ f^{-1}$. For all $v \in \mathcal{V}$ we get

$$L(p) = \int_K p(x)\ \mathrm{d}\mu(x) = \int_K p(x)\ \mathrm{d}(\lambda \circ f^{-1})(x) \overset{\text{Lemma 13.1.1}}{=} \int_0^1 p(f(t))\ \mathrm{d}\lambda(t). \tag{56}$$

(ii)→(i): By (56) $\mu = \lambda \circ f^{-1}$ is a representing measure of L, i.e., L is a K-moment functional. To show that $\operatorname{supp}\mu = K$ holds, let $U \subseteq K$ be open. Since f is continuous, $f^{-1}(U) \subseteq [0,1]$ is open and therefore $\mu(U) = \lambda(f^{-1}(U)) > 0$. □

So far we transformed moment functionals to $[0,1]$-moment functionals. We have seen that e.g. $\mathbb{R}^n$-moment functionals can not be continuously transformed into $[0,1]$-moment functionals. But we can transform $\mathbb{R}^n$-moment functionals continuously into $[0,\infty)$-moment functionals. We need the following.

Lemma 13.2.15. *Let $n \in \mathbb{N}$ and $\varepsilon > 0$. Then there exists a continuous and surjective function $f_\varepsilon : [0,\infty) \to \mathbb{R}^n$ with*

$$t - \varepsilon \leq \|f_\varepsilon(t)\| \leq t + \varepsilon$$

for all $t \geq 0$ and there exists a measurable function $g_\varepsilon : \mathbb{R}^n \to [0,\infty)$ such that

$$f_\varepsilon(g_\varepsilon(x)) = x \qquad \text{and} \qquad \|x\| - \varepsilon \leq g_\varepsilon(x) \leq \|x\| + \varepsilon$$

for all $x \in \mathbb{R}^n$.

Proof. Set

$$A_n := \{x \in \mathbb{R}^n \,|\, (n-1)\cdot\varepsilon \leq \|x\| \leq n\cdot\varepsilon\}$$

for all $n \in \mathbb{N}$. Then all A_n's are compact and path-connected and by the Hahn–Mazurkiewicz' Theorem 13.1.6 there exist continuous and surjective functions $f_{\varepsilon,n} : [(n-1)\cdot\varepsilon, n\cdot\varepsilon] \to A_n$ for all $n \in \mathbb{N}$ such that $f_{\varepsilon,n}(n\cdot\varepsilon) = f_{\varepsilon,n+1}(n\cdot\varepsilon)$, i.e., $\|f_{\varepsilon,n}(n\cdot\varepsilon)\| = \|f_{\varepsilon,n+1}(n\cdot\varepsilon)\| = n\cdot\varepsilon$ for all $n \in \mathbb{N}$. Since $\mathbb{R}^n = \bigcup_{n\in\mathbb{N}} A_n$ define $f_\varepsilon : [0,\infty) \to \mathbb{R}^n$ by $f_\varepsilon|_{[n-1,n]} := f_{\varepsilon,n}$. Then for $t \in [(n-1)\cdot\varepsilon, n\cdot\varepsilon]$ we have

$$t - \varepsilon \leq (n-1)\cdot\varepsilon \leq \|f_\varepsilon(t)\| = \|f_{\varepsilon,n}(t)\| \leq n\cdot\varepsilon \leq t+\varepsilon. \tag{57}$$

Since $f : [0,\infty) \to \mathbb{R}^n$ is surjective and $[0,\infty)$ and $\mathbb{R}^n$ are Souslin sets by Lemma 13.1.7 then by Jankoff's Theorem 13.1.12 there exists a $g_\varepsilon : \mathbb{R}^n \to [0,\infty)$ with $f_\varepsilon(g_\varepsilon(x)) = x$ for all $x \in \mathbb{R}^n$. (57) implies

$$g_\varepsilon(x) - \varepsilon \leq \|x\| = \|f_\varepsilon(g_\varepsilon(x))\| \leq g_\varepsilon(x) + \varepsilon$$

and therefore $\|x\| - \varepsilon \leq g_\varepsilon(x) \leq \|x\| + \varepsilon$ for all $x \in \mathbb{R}^n$. □

Theorem 13.2.16. *Let $n \in \mathbb{N}$, $f : [0,\infty) \to \mathbb{R}^n$ be a continuous and surjective function, and $\mathcal{V}$ be a vector space of measurable functions on $\mathbb{R}^n$. Then for all linear functionals $L : \mathcal{V} \to \mathbb{R}$ the following are equivalent:*

(i) $L : \mathcal{V} \to \mathbb{R}$ is a moment functional.

(ii) $\tilde{L} : \mathcal{V}\circ f \to \mathbb{R}$ defined by $\tilde{L}(v\circ f) := L(v)$ is a $[0,\infty)$-moment functional.

I.e., $L \overset{sc}{\rightsquigarrow} \tilde{L}$. If $\tilde{\mu}$ is a representing measure of $\tilde{L}$, then $\tilde{\mu}\circ f^{-1}$. There exists a function $g : \mathbb{R}^n \to [0,\infty)$ such that $f(g(x)) = x$ for all $x \in \mathbb{R}^n$ and if μ is a representing measure of L, then $\mu \circ g^{-1}$ is a representing measure of $\tilde{L}$.

Proof. Since $\mathbb{R}^n$ and $[0,\infty)$ are Souslin sets and f is surjective, by Jankoff's Theorem 13.1.12 there exists a function $g:\mathbb{R}^n\to[0,\infty)$ such that $f(g(x)) = x$ for all $x\in\mathbb{R}^n$. It follows that $\tilde{L}$ is well defined by $\tilde{L}(v\circ f) = L(v)$.

(i)→(ii): Let μ be a representing measure of L, then

$$\begin{aligned}\tilde{L}(v\circ f) = L(v) = \int_{\mathbb{R}^n} v(x)\ \mathrm{d}\mu(x) &= \int_{\mathbb{R}^n} v(f(g(x)))\ \mathrm{d}\mu(x)\\ &\overset{\text{Lemma 13.1.1}}{=} \int_0^\infty (v\circ f)(t)\ \mathrm{d}(\mu\circ g^{-1})(t),\end{aligned}$$

i.e., $\mu\circ g^{-1}$ is a representing measure of $\tilde{L}$.

(ii)→(i): Let $\tilde{\mu}$ be a representing measure of $\tilde{L}$, then

$$L(v) = \tilde{L}(v\circ f) = \int_0^\infty (v\circ f)(t)\ \mathrm{d}\tilde{\mu}(t) \overset{\text{Lemma 13.1.1}}{=} \int_{\mathbb{R}^n} v(x)\ \mathrm{d}(\tilde{\mu}\circ f^{-1})(x),$$

i.e., $\tilde{\mu}\circ f^{-1}$ is a representing measure of L. □

THE END

A. Appendix

A.1. Hahn–Banach[31] dominated Extension Theorem

Hahn–Banach dominated Extension Theorem A.1.1. *Let $\mathcal{V}$ be a real linear space, $q : \mathcal{V} \to \mathbb{R}$ be a sublinear functional, i.e.,*

$$q(x+y) \leq q(x) + q(y) \qquad \text{and} \qquad q(t\cdot x) = t\cdot q(x)$$

for all $x, y \in \mathcal{V}$ and $t \geq 0$, let $\mathcal{W} \subseteq \mathcal{V}$ be a subspace, and $L : \mathcal{W} \to \mathbb{R}$ be a linear functional such that $L(x) \leq q(x)$ for all $x \in \mathcal{W}$. Then there exists an extension $\tilde{L} : \mathcal{V} \to \mathbb{R}$ of L such that $\tilde{L}(x) \leq q(x)$ for all $x \in \mathcal{V}$ holds.

A.2. The Riesz–Markov–Kakutani Representation Theorem

Riesz–Markov–Kakutani Representation Theorem A.2.1 ([Rie09, Mar38, Kak41])**.** *Let $\mathcal{X}$ be a locally compact Hausdorff space and $L : C_c(\mathcal{X}, \mathbb{R}) \to \mathbb{R}$ be a linear functional such that $L(f) \geq 0$ for all $f \in C_c(\mathcal{X}, \mathbb{R})_+$. Then there exists a unique $\mu \in \mathcal{M}(\mathcal{X})$ with*

$$L(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$$

for all $f \in C_c(\mathcal{X}, \mathbb{R})$.

The present representation theorem was developed in several stages. A first version for continuous functions on the unit interval $[0,1]$ is by Frigyes Riesz[32] [Rie09]. It was extended by Andrey Markov[33] to some non-compact spaces [Mar38] and then by Shizuo Kakutani[34] to locally compact Hausdorff spaces [Kak41]. Interestingly, it already follows from Daniell's[35] Representation Theorem A.6.2 [Dan18, Dan20] with Urysohn's Lemma[36] [Ury25].

A.3. Stone–Weierstraß[37] Theorem

Stone–Weierstraß Theorem A.3.1. *Let $\mathcal{X}$ be a compact Hausdorff space and $A \subseteq C(\mathcal{X}, \mathbb{R})$ be a unital algebra. Then A is dense in $C(\mathcal{X}, \mathbb{R})$ if and only if A separates points.*

[31] Hans Hahn (27 September 1879, Vienna – 24 July 1934, Vienna);
Stefan Banach (30 March 1892, Krakow – 31 August 1945, Lemberg)
[32] Frigyes Riesz (22 January 1880, Györ (Hungary) – 28 February 1956, Budapest)
[33] Andrey Andreyevich Markov (14 June 1856, Rjasan (Russia) – 20 July 1922, Petrograd)
[34] Shizuo Kakutani (28 August 1911, Osaka – 17 August 2004, New Haven (Connecticut))
[35] Percy John Daniell (9 January 1889, Valparaíso (Chile) – 25 May 1946, Sheffield (UK))
[36] Pavel Samuilovich Urysohn (3 February 1898, Odessa – 17 August 1924, Batz-sur-Mer (France))
[37] Marshall Harvey Stone (8 April 1903, New York City – 9 January 1989, Madras (India));
Karl Theodor Wilhelm Weierstraß (31 October 1815, Ostenfelde (near Ennigerloh) – 19 February 1897, Berlin)

A.4. Carathéodory's Theorem

Carathéodory's Theorem A.4.1 (conic version, see e.g. [Roc72, Cor. 17.1.2]). *Let $d \in \mathbb{N}$, $\{C_i \,|\, i \in I\}$ be an arbitrary collection of non-empty convex sets in $\mathbb{R}^d$, and let K be the convex cone generated by the union of the collection. Then every non-zero vector of K can be expressed as a non-negative linear combination of d or fewer linearly independent vectors, each belonging to a different C_i.*

A.5. Sard's[38] Theorem

Definition A.5.1. Let $n, m \in \mathbb{N}$, $\mathcal{X} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{X} \to \mathbb{R}^m$ be a C^1-mapping. $x \in \mathcal{X}$ is called a *regular* point if $Df(x)$ has full rank. Otherwise $x \in \mathcal{X}$ is called *singular*. A point $y \in \mathbb{R}^m$ is called a *regular value* if $f^{-1}(y)$ is empty or consists solely of regular points. Otherwise $y \in \mathbb{R}^m$ is called a *singular value*.

Sard's Theorem A.5.2 ([Sar42]). *Let $n, m \in \mathbb{N}$, $\mathcal{X} \subseteq \mathbb{R}^n$ be open, and $f : \mathcal{X} \to \mathbb{R}^m$ be a C^r-mapping with $r > \max\{0, n - m\}$. Then the set of singular values of f has m-dimensional Lebesgue measure zero and the regular values are dense in $\mathbb{R}^m$.*

There is also an algebraic version of Sard's Theorem, see e.g. [BCR98].

A.6. Daniell's Representation Theorem

Definition A.6.1. Let $\mathcal{X}$ be a space. We call a set $\mathcal{F}$ of functions $f : \mathcal{X} \to \mathbb{R}$ a *lattice* (*of functions*) if the following holds:

i) $c \cdot f \in \mathcal{F}$ for all $c \geq 0$ and $f \in \mathcal{F}$,

ii) $f + g \in \mathcal{F}$ for all $f, g \in \mathcal{F}$,

iii) $\inf(f, g) \in \mathcal{F}$ for all $f, g \in \mathcal{F}$,

iv) $\inf(f, c) \in \mathcal{F}$ for all $c \geq 0$ and $f \in \mathcal{F}$, and

v) $g - f \in \mathcal{F}$ for all $f, g \in \mathcal{F}$ with $f \leq g$.

Daniell's Representation Theorem A.6.2 (P. J. Daniell 1918 [Dan18]). *Let $\mathcal{F}$ be a lattice of functions on a space $\mathcal{X}$ and let $L : \mathcal{F} \to \mathbb{R}$ be such that*

i) $L(f + g) = L(f) + L(g)$ *for all* $f, g \in \mathcal{F}$,

ii) $L(c \cdot f) = c \cdot L(f)$ *for all* $c \geq 0$ *and* $f \in \mathcal{F}$,

iii) $L(f) \leq L(g)$ *for all* $f, g \in \mathcal{F}$ *with* $f \leq g$,

iv) $L(f_n) \nearrow L(g)$ *as* $n \to \infty$ *for all* $g \in \mathcal{F}$ *and* $f_n \in \mathcal{F}$ *with* $f_n \nearrow g$.

[38] Arthur Sard (28 July 1909, New York City – 31 August 1980, Basel)

Then there exists a measure μ on $(\mathcal{X},\mathcal{A})$ with

$$\mathcal{A} := \sigma(\{f^{-1}((-\infty,a]) \,|\, a\in\mathbb{R},\ f\in\mathcal{F}\})$$

such that

$$L(f) = \int_{\mathcal{X}} f(x)\ \mathrm{d}\mu(x)$$

for all $f\in\mathcal{F}$.

The most impressive part is that the functional $L:\mathcal{F}\to\mathbb{R}$ lives only on a lattice $\mathcal{F}$ of functions $f:\mathcal{X}\to\mathbb{R}$ where $\mathcal{X}$ is a set without any structure. Daniell's Representation Theorem A.6.2 provides a representing measure μ including the σ-algebra $\mathcal{A}$ of the measurable space $(\mathcal{X},\mathcal{A})$.

Riesz–Markov–Kakutani Representation Theorem A.2.1 follows directly from Daniell's Representation Theorem A.6.2. $C_0(\mathcal{X},\mathbb{R})$, $\mathcal{X}$ a locally compact Hausdorff space, is a lattice of functions, (i) and (ii) are the linearity of L, (iii) non-negativity of L, and the continuity condition (iv) of L follows easily from uniform convergence in $C_0(\mathcal{X},\mathbb{R})$.

Index

A

adapted . . . 6
apolar
  scalar product . . . 35

B

Banach
  Stefan . . . 83
bases
  dual . . . 36
basic
  set of nodes . . . 36
Borel
  map . . . 70

C

$\mathcal{C}_{\mathsf{A}}$ . . . 21
Carathéodory
  Constantin . . . 20
  number . . . 21
$\mathcal{C}_{\mathsf{A}}(s)$ . . . 21
Choi–Lam
  polynomial . . . 15
Choquet
  Gustave Alfred Arthur . . . 6
  theory . . . 6
$\mathcal{C}(L)$ . . . 21
cone
  dual . . . 17
  moment . . . 17
  normal . . . 17
Cubature formula
  Gaussian . . . 19
Curto
  Raúl Enrique . . . 30
curve
  moment . . . 17
  Peano . . . 70
  space filling . . . 24
    Schönberg . . . 25

D

Daniell
  Percy John . . . 83
  Representation Theorem . . . 83, 84
decomposition
  Waring . . . 38
Δ . . . 40
determinate . . . 5, 13
  K-determinate . . . 5
dominate . . . 6
dual
  bases . . . 36
dual cone . . . 17

E

eigenvalue
  generalized . . . 33
  problem
    generalized . . . 33
eigenvector
  generalized . . . 33
E_+ . . . 5
exposed face . . . 17
extension
  flat . . . 30

F

face . . . 17
  exposed . . . 17
$f_{\text{Berg−Christensen−Jensen}}$ . . . 15
f_{Harris} . . . 15
Fialkow
  Lawrance . . . 30
flat . . . 30
  extension . . . 30
  linear functional . . . 30
f_{Motzkin} . . . 14
f_{Robinson} . . . 15
$f_{\text{Schmüdgen}}$ . . . 15
function
  Hilbert . . . 25
  Schwartz . . . 51

functional
  K-moment 5
  linear
    C-positive 5
    derivative 43
    strictly C-positive 5
  moment 5
  Riesz 9
  sublinear 83

G
Gaussian
  cubature formula 19
Gauß
  Carl Friedrich 19
generalized
  eigenvalue 33
    problem 33
  eigenvector 33

H
Hahn
  Hans 83
halfspace 17
  containing 17
Hamburger
  Hans Ludwig 11
  moment problem 11
Hankel
  Hermann 9
  matrix 9
Harris
  polynomial 15
Hausdorff
  Felix 13
  moment problem 12
Haviland
  Edward Kenneth 8
  Theorem 8
$H_d(s)$ 9
HF_R 25
Hilbert
  David 14
  function 25
  Nullstellensatz 27
  Theorem 14
$h_{n,2d}$ 40
$\mathcal{H}_{n,d}$ 35
homogenization 26
hyperplane 17
  supporting 17

I
identity
  Marsden 37
isomorphic
  mod0 71
isomorphism 71
  point 71

K
Kakutani
  Shizuo 83

L
Laplace operator 40
lattice
  of functions 84
Lemma
  Urysohn 83
linear form
  power 35
L_μ 5
L_s 9

M
$\mathcal{M}_{\mathsf{A}}$ 16
map
  Borel 70
Markov
  Andrey Andreyevich 83
Marsden
  identity 37
matrix
  Hankel 9
measure
  A-measurable 16
  derivative 44
  representing 5

set of ... 5
mod0 ... 71
Möller
H. Micheal ... 24
lower bound ... 24
moment ... 9
cone ... 17
curve ... 17
derivative ... 43
map ... 21
problem ... 3
Hamburger ... 11
Hausdorff ... 12
Stieltjes ... 11
Motzkin
polynomial ... 14
Theodore Samuel ... 14
$\mathcal{M}(\mathcal{X})$ ... 4

N

$\mathsf{N}_{n,d}$ ... 35
nodes
basic set of ... 36
$\mathrm{Nor}_{\mathsf{A}}(s)$ ... 17
normal
vector ... 17
normal cone ... 17
Nullstellensatz
Hilbert ... 27

P

Peano
curve ... 70
point
isomorphism ... 71
regular ... 84
singular ... 84
polynomial
Berg–Christensen–Jensen ... 15
Choi–Lam ... 15
Harris ... 15
Motzkin ... 14
non-negative ... 8
Robinson ... 15
Schmüdgen ... 15
Positivstellensatz
Schmüdgen ... 16
$\mathrm{Pos}(K)$ ... 8
$\mathrm{Pos}(n,d)$ ... 14
power
linear form ... 35

R

rank
Waring ... 41
regular
point ... 84
sequence ... 26
value ... 84
Richter
Hans Werner ... 18
Theorem ... 18
Riesz
Frigyer ... 83
functional ... 9
Robinson
polynomial ... 15
Rogosinski
W. W. ... 20
Rosenbloom
P. C. ... 19

S

$\mathcal{S}_{\mathsf{A}}$ ... 17
s_{A} ... 17
$S_{\mathsf{A},k}$ ... 21
$\mathcal{S}_{\mathsf{A},k}$ ... 21
Sard
Arthur ... 84
Theorem ... 84
scalar product
apolar ... 35
standard ... 35
Schmüdgen
Konrad ... 15
polynomial ... 15
Positivstellensatz ... 16
Theorem ... 16

Schwartz function 51
Schönberg
  Isaac Jacob 25
  space filling curve 25
sequence
  K-moment 9
  moment 9
    Hamburger 11
    Hausdorff 13
    Stieltjes 12
  positive definite 10
  positive semidefinite 10
  regular 26
set
  Souslin 69
set of nodes
  basic 36
shift 10
$\Sigma(n,d)$ 14
$\sum \mathbb{R}[x_1,\dots,x_n]^2$ 8
singular
  point 84
  value 84
Souslin
  set 69
  space 69
space
  adapted 6
  filling curve 24
    Schöneberg 25
  Lebesgue–Rohlin 72
  Souslin 69
$\mathcal{S}(\mathbb{R}^n)$ 51
Stieltjes
  moment problem 11
  Thomas Joannes 12
Stone
  Marshall Harvey 83
sums of squares 8
support 17
supporting
  hyperplane 17

T

Tchakaloff
  M. V. 19
tensor
  elementary 42
  space 42
  symmetric 42
Theorem
  Alexander–Hirschowitz 24
  Basic Representation 7
  Daniell's Representation 83, 84
  $d-1$ 22
  Flat Extension 30
  Hahn–Banach 83
  Hahn–Mazurkiewicz 70
  Hamburger 11
  Hausdorff 13
  Haviland 8
  Hilbert 14
  Jankoff 71
  Kolesnikov 72
  Richter 18
  Riesz–Markov–Kakutani
    Representation 83
  Sard 84
  Schmüdgen 16
    Positivstellensatz 16
  Stieltjes 12
  Stone–Weierstraß 83
  Urysohn 83
transformation $\rightsquigarrow$ 73
  continuous $\overset{c}{\rightsquigarrow}$ 73
  strong $\overset{s}{\rightsquigarrow}$ 73
  strong and continuous $\overset{sc}{\rightsquigarrow}$ 73

U

Urysohn
  Lemma 83
  Pavel Samuilovich 83

V

value
  regular 84
  singular 84

vector
  normal ... 17

W

Wald
  A. ... 19
Waring
  decomposition ... 38
  Edward ... 35
  rank ... 41
Weierstraß
  Karl Theodor Wilhelm ... 83
$\mathcal{W}(f)$ ... 41
$\mathcal{W}(n,2d)$ ... 41

X

X^γ ... 10

Z

zero
  set ... 18
$\mathcal{Z}(f)$ ... 18

References

[AFS16] C. Améndola, J.-C. Faugère, and B. Sturmfels, *Moment varieties of gaussian mixtures*, J. Alg. Stat. **7** (2016), 14–28.

[AH95] J. Alexander and A. Hirschowitz, *Polynomial interpolation in several variables*, J. Alg. Geom. **4** (1995), 201–222.

[AK62] N. I. Ahiezer and M. Kreĭn, *Some Questions in the Theory of Moments*, American Mathematical Society, Providence, Rhode Island, 1962.

[Akh65] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*, Oliver & Boyd, Edinburgh, 1965.

[Alo99] N. Alon, *Combinatorial Nullstellensatz*, Combin. Probab. Comput. **8** (1999), 7–29.

[Ana06] G. A. Anastassiou, *Applications of geometric moment theory related to optimal portfolio management*, Comput. Math. Appl. **51** (2006), 1405–1430.

[Art27] E. Artin, *Über die Zerlegung definiter Funktionen in Quadrate*, Abh. math. Sem. Hamburg **5** (1927), 100–115.

[Bal61] M. L. Balinski, *An algorithm for finding all vertices of convex polyhedral sets*, J. Soc. Indust. Appl. Math. **9** (1961), 72–88.

[Bar91] A. I. Barvinok, *Calculation of exponential integrals*, Zap. Nauč. Semin. POMI **192** (1991), 175–176.

[Bar92] ______, *Exponential integrals and sums over convex polyhedra*, Funkc. Anal. Prilozh. **26** (1992), 64–66.

[BCJ79] C. Berg, J. P. R. Christensen, and C. U. Jensen, *A remark on the multidimensional moment problem*, Math. Ann. **243** (1979), 163–169.

[BCR98] J. Bochnak, M. Coste, and M.-F. Roy, *Real Algebraic Geometry*, Springer-Verlag, Berlin, 1998.

[BGL07] B. Beckermann, G. H. Golub, and G. Labahn, *On the numerical condition of a generalized Hankel eigenvalue problem*, Numer. Math. **106** (2007), 41–68.

[Bog07] V. I. Bogachev, *Measure Theory*, Springer-Verlag, Berlin, 2007.

[Bri88] M. Brion, *Points entiers dans les polyèdres convexes*, Ann. Sci. Éc. Norm. Super. **21** (1988), 653–663.

[BT06] C. Bayer and J. Teichmann, *The proof of Tchakaloff's theorem*, Proc. Amer. Math. Soc. **134(10)** (2006), 3035–3040.

[CdD22] R. Curto and P. J. di Dio, *Time-dependent moments from the heat equation and a transport equation*, Int. Math. Res. Notices (2022), https://doi.org/10.1093/imrn/rnac244.

[CF96] R. Curto and L. A. Fialkow, *Solution of the truncated moment problem for flat data*, Mem. Amer. Math. Soc. **119** (1996), no. 568.

[CF98] ______, *Flat extensions of positive moment matrices: recursively generated relations*, Mem. Amer. Math. Soc. **136** (1998), no. 648.

[CKP20] D. Cifuentes, T. Kahle, and P. Parrilo, *Sums of squares in Macaulay2*, J. Softw. Algebra Geom. **10** (2020), 17–24.

[CL77] M. D. Choi and T.-Y. Lam, *Extremal positive semi-definite forms*, Math. Ann. **231** (1977), 1–18.

[Dan18] P. J. Daniell, *A General Form of Integral*, Ann. Math. **19** (1918), 279–294.

[Dan20] ______, *Further Properties of the General Integral*, Ann. Math. **21** (1920), 203–220.

[dD19] P. J. di Dio, *The multidimensional truncated Moment Problem: Gaussian and Log-Normal Mixtures, their Carathéodory Numbers, and Set of Atoms*, Proc. Amer. Math. Soc. **147** (2019), 3021–3038.

[dD23a] ______, *The multidimensional truncated moment problem: Gaussian mixture reconstruction from derivatives of moments*, J. Math. Anal. Appl. **517** (2023), 126592.

[dD23b] ______, *Transformations of Moment Functionals*, Integr. Equ. Oper. Theory **95** (2023), 2, 29 pages.

[dDK21] P. J. di Dio and M. Kummer, *The multidimensional truncated Moment Problem: Carathéodory Numbers from Hilbert Functions*, Math. Ann. **380** (2021), 267–291.

[dDS18] P. J. di Dio and K. Schmüdgen, *The multidimensional truncated Moment Problem: Carathéodory Numbers*, J. Math. Anal. Appl. **461** (2018), 1606–1638.

[dDS22] P. J. di Dio and K. Schmüdgen, *The multidimensional truncated moment problem: The moment cone*, J. Math. Anal. Appl. **511** (2022), 126066, 38 pages.

[Eva10] L. C. Evans, *Partial Differential Equations*, 2 ed., Amercian Mathematical Society, Providence, Rhode Island, 2010.

[Fed69] H. Federer, *Geometric Measure Theory*, Springer-Verlag, Berlin, 1969.

[FP01] C. F. Floudas and P. M. Pardalos (eds.), *Encyclopedia of Optimization*, vol. 1, Kluwer Academic Publishers, Dordrecht, 2001.

[Gau15] C. F. Gauß, *Methodus nova integralium valores per approximationem inveniendi*, Comm. Soc. Sci. Göttingen Math. **3** (1815), 29–76.

[GLPR12] N. Gravin, J. Lasserre, D. V. Pasechnik, and S. Robins, *The inverse moment problem for convex polytopes*, Discrete Comput. Geom. **48** (2012), 596–621.

[GMV99] G. H. Golub, P. Milfar, and J. Varah, *A stable numberical method for inverting shape from moments*, SIAM J. Sci. Comput. **21** (1999), no. 4, 1222–1243.

[GNPR14] N. Gravin, D. Nguyen, D. V. Pasechnik, and S. Robins, *The inverse moment problem for convex polytopes: Implementation aspects*, arXiv:1409.3130v2.

[GPSS18] N. Gravin, D. Pasechnik, B. Shapiro, and M. Shapiro, *On moments of a polytope*, Anal. Math. Phys. **8** (2018), 255–287.

[Gru09] G. Grubb, *Distributions and operators*, Spinger, New York, 2009.

[GS] D. R. Grayson and M. E. Stillman, *Macaulay2, a software system for research in algebraic geometry*, Available at `http://www.math.uiuc.edu/Macaulay2/`.

[Hah14] H. Hahn, *Mengentheoretische Charakterisierung der stetigen Kurven*, Sitzungsber. Akad. Wiss. Wien **123** (1914), 2433–2489.

[Ham20] H. L. Hamburger, *Über eine Erweiterung des Stieltjesschen Momentenproblems*, Math. Ann. **81** (1920), 235–319.

[Har99] W. R. Harris, *Real Even Symmetric Ternary Forms*, J. Alg. **222** (1999), 204–245.

[Hau21] F. Hausdorff, *Summationsmethoden und Momentenfolgen I*, Math. Z. **9** (1921), 74–109.

[Hav35] E. K. Haviland, *On the momentum problem for distribution functions in more than one dimension*, Amer. J. Math. **57** (1935), 562–572.

[Hav36] ______, *On the momentum problem for distribution functions in more than one dimension II*, Amer. J. Math. **58** (1936), 164–168.

[Hil88] D. Hilbert, *Über die Darstellung definiter Formen als Summe von Formenquadraten*, Math. Ann. **32** (1888), 342–350.

[Jan41] W. Jankoff, *Sur l'uniformisation des ensembles A*, Dokl. Akad. Nauk SSSR **30** (1941), 591–592, French transl.: C. R. (Doklady) Acad. Sci. URSS **30** (1941), 597–598.

[Kak41] S. Kakutani, *Concrete representation of abstract (M)-spaces (A characterization of the space of continuous functions)*, Ann. Math. Series 2 **42** (1941), 994–1024.

[Kem68] J. H. B. Kemperman, *The General Moment Problem, a Geometric Approach*, Ann. Math. Stat. **39** (1968), 93–122.

[Kem71] ______, *Moment problems with convexity conditions I*, Optimizing Methods in Statistics (J. S. Rustagi, ed.), Acad. Press, 1971, pp. 115–178.

[Kem87] ______, *Geometry of the Moment Problem*, Proc. Sym. Appl. Math. **37** (1987), 16–53.

[KN77] M. G. Kreĭn and A. A. Nudel'man, *The Markow Moment Problem and Extremal Problems*, American Mathematical Society, Providence, Rhode Island, 1977, translation of the Russian original from 1973.

[Kol99] A. V. Kolesnikov, *On continuous images of the Lebesgue measure*, Matem. Zametki. **65** (1999), 790–793, English transl.: Math. Notes **65** (1999), 665–668.

[KSS18] K. Kohn, B. Shapiro, and B. Sturmfels, *Moment varieties of measures on polytopes*, arXiv:1807.10258v1.

[Las15] J.-B. Lasserre, *An introduction to polynomial and semi-algebraic optimization*, Cambridge University Press, Cambridge, 2015.

[Lau09] M. Laurent, *Sums of squares, moment matrices and optimization over polynomials*, Emerging application of algebraic geometry, IMA Vol. Math. Appl., vol. 149, Springer, New York, 2009, pp. 157–270.

[Law91] J. Lawrence, *Polytope volume computation*, Math. Comput. **57** (1991), 259–271.

[LL01] E. H. Lieb and M. Loss, *Analysis*, Graduate Texts in Mathematics, vol. 14, Amercian Mathematical Society, Providence, Rhode Island, 2001.

[LR82] Y. T. Lee and A. A. G. Requicha, *Algorithms for computing the volume and other integral properties of solids. I. known methods and open issues*, Comm. ACM **25** (1982), 635–641.

[Mar38] A. Markov, *On mean values and exterior densities*, Rec. Math. Moscou. N.S. **4** (1938), 165–190.

[Mar08] M. Marshall, *Positive Polynomials and Sums of Squares*, Mathematical Surveys and Monographs, no. 146, American Mathematical Society, Rhode Island, 2008.

[Maz20] S. Mazurkiewicz, *Sur les lignes de Jordan*, Fund. Math. **1** (1920), 166–209.

[MN68] M. Maňas and J. Nedoma, *Finding all vertices of a convex polyhedron*, Numer. Math. **12** (1968), 226–229.

[Möl76] H. M. Möller, *Kubaturformeln mit minimaler Knotenzahl*, Numer. Math. **25** (1976), 185–200.

[Mot67] T. S. Motzkin, *The arithmetic-geometric inequality*, Inequalities (New York) (O. Shisha, ed.), Proc. of Sympos. at Wright-Patterson AFB, August 19–27, 1965, Academic Press, 1967, pp. 205–224.

[MR80] T. H. Matheiss and D. S. Rubin, *A survey and comparison of methods for finding all vertices of convex polyhedral sets*, Math. Oper. Res. **5** (1980), 167–185.

[Mün14] C. H. Müntz, *Mathematische Abhandlungen Hermann Amandus Schwarz zu seinem fünfzigjährigen Doktorjubiläum am 6. August 1914 gewidmet von Freunden und Schülern.*, ch. Über den Approximationssatz von Weierstrass, pp. 303–312, Springer, Berlin, 1914.

[MVKW95] P. Milanfar, G. Verghese, W. Karl, and A. Willsky, *Reconstructing polygons from moments with connections to array processing*, IEEE Trans. Signal Proc. **43** (1995), 432–443.

[Pfi67] A. Pfister, *Zur Darstellung definiter Funktionen als Summe von Quadraten*, Invent. Math. **4** (1967), 229–237.

[PK92] A. V. Pukhlikov and A. G. Khovanskii, *The Riemann–Roch theorem for integrals and sums of quasipolynomials on virtual polytopes*, Algebra Anal. **4** (1992), 188–216.

[Ric57] H. Richter, *Parameterfreie Abschätzung und Realisierung von Erwartungswerten*, Bl. Deutsch. Ges. Versicherungsmath. **3** (1957), 147–161.

[Rie09] F. Riesz, *Sur les opérations fonctionnelles linéaires*, C. R. Acad. Sci. Paris **144** (1909), 1409–1411.

[Rob69] R. M. Robinson, *Some definite polynomials which are not sums of squares of real polynomials*, Notices Amer. Math. Soc. **16** (1969), 554.

[Roc72] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, New Jersey, 1972.

[Rog58] W. W. Rogosinski, *Moments of non-negative mass*, Proc. R. Soc. Lond. A **245** (1958), 1–27.

[Ros52] P. C. Rosenbloom, *Quelques classes de problème extrémaux. II*, Bull. Soc. Math. France **80** (1952), 183–215.

[RS18] C. Riener and M. 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. Compl. **45** (2018), 22–54.

[Sag94] H. Sagan, *Space-Filling Curves*, Springer-Verlag, 1994.

[Sar42] A. Sard, *The measure of critical values of differentiable maps*, Bull. Amer. Math. Soc. **48** (1942), 883–890.

[Sch79] K. Schmüdgen, *A positive polynomial which is not a sum of squares. A positive, but not strongly positive functional*, Math. Nachr. **88** (1979), 385–390.

[Sch91] ———, *The K-moment problem for closed semi-algebraic sets*, Math. Ann. **289** (1991), 203–206.

[Sch14] R. Schneider, *Convex Bodies: The Brunn–Minkowski Theory*, Cambridge University Press, Cambridge, 2014.

[Sch17] K. Schmüdgen, *The Moment Problem*, Springer, New York, 2017.

[Sim11] B. Simon, *Convexity: An Analytic Viewpoint*, Cambridge University Press, Cambridge, 2011.

[Sol15] V. Soltan, *Lectures on Convex Sets*, World Scientific, Hackensack, NJ, 2015.

[Sti94] T. J. Stieltjes, *Recherches sur les fractions continues*, Ann. Fac. Sci. Toulouse **8** (1894), no. 4, J1–J122.

[Szá16] O. Szász, *Über die Approximation stetiger Funktionen durch lineare Aggregate von Potenzen*, Math. Ann. **77** (1916), 482–496.

[Tch57] M. V. Tchakaloff, *Formules de cubatures méchaniques a coefficients non négatifs*, Bull. Sci. Math. **81** (1957), 123–134.

[Ury25] P. Urysohn, *Über die Mächtigkeit der zusammenhängenden Mengen*, Math. Ann. **94** (1925), 262–295.

[Wal39] A. Wald, *Limits of distribution function determined by absolute moment and inequalities satisfied by absolute moments*, Trans. Amer. Math. Soc. **46(2)** (1939), 280–306.