The classical moment problem

Sasha Sodin*

March 5, 2019

> These lecture notes were prepared for a 10-hour introductory mini-course at LTCC. A cap was imposed at 50 pages; even thus, I did not have time to cover the sections indicated by an asterisk. If you find mistakes or misprints, please let me know. Thank you very much!

The classical moment problem studies the map $\mathcal{S}$ taking a (positive Borel) measure μ on $\mathbb{R}$ to its moment sequence $(s_k)_{k\geq 0}$,

$$s_k[\mu] = \int \lambda^k d\mu(\lambda) \ .$$

The map is defined on the set of measures with finite moments:

$$\mu(\mathbb{R} \setminus [-R, R]) = O(R^{-\infty}) \ , \quad \text{i.e.} \quad \forall k \ \ \mu(\mathbb{R} \setminus [-R, R]) = O(R^{-k}) \ .$$

The two basic questions are

1. existence: characterise the image of $\mathcal{S}$, i.e. for which sequences $(s_k)_{k\geq 0}$ of real numbers can one find μ such that $s_k[\mu] = s_k$ for $k = 0, 1, 2, \cdots$?
2. uniqueness, or determinacy: which sequences in the image have a unique pre-image, i.e. which measures are characterised by their moments? In the case of non-uniqueness, one may wish to describe the set of all solutions.

The classical moment problem originated in the 1880-s, and reached a definitive state by the end of the 1930-s. One of the original sources of motivation came from probability theory, where it is important to have verifiable sufficient conditions for determinacy. Determinacy is also closely related to several problems in classical analysis, particularly, to the study of the map taking a (germ of a) smooth function f to the sequence of its Taylor coefficients $(f^{(k)}(0))_{k\geq 0}$. Existence in the moment problem is a prototype of the problem of extension of a positive functional, and it gave the impetus for the development of several functional-analytic tools.

*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom. E-mail: a.sodin@qmul.ac.uk

The moment problem also enjoys a mutual relation with the spectral theory of self-adjoint operators. In fact, the spectral theorem for bounded self-adjoint operators can be deduced from the existence theorem for the moment problem. Further, the moment problem provides simple and yet non-trivial examples of various notions from the abstract theory of unbounded symmetric and self-adjoint operators.

The classical monograph Akhiezer [1965] is still the best reference on the moment problem and related topics. We touch only briefly on the approach (originating in the work of Chebyshev) to the moment problem as an extremal problem; see Krein and Nudel′man [1977]. The classical reference on quasianalyticity is Carleman [1926].

1 Introduction

1.1 A motivating example

The following problem is a variant of the one considered by Pafnuty Chebyshev in the 1880-s[1]. Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures. Assume:

$$\forall k \geq 1 \qquad \lim_{n\to\infty} s_k[\mu_n] = \begin{cases} \frac{k!}{2^{k/2}(k/2)!} \ , & k \text{ is even} \\ 0 \ , & k \text{ is odd} \end{cases} \tag{1.1}$$

Example 1.1. *Let $X_1, X_2, X_3, \cdots$ be independent, identically distributed random variables with $\mathbb{E}X_1 = 0$, $\mathbb{E}X_1^2 = 1$ and $\mathbb{E}|X_1|^k < \infty$ for any $k \geq 3$. Let μ_n be the law of $\frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n)$, i.e.*

$$\mu_n(B) = \mathbb{P}\left\{ \frac{1}{\sqrt{n}}(X_1 + X_2 + \cdots + X_n) \in B \right\} .$$

Then (1.1) holds.

Exercise 1.2. Prove this.

Exercise 1.3. Prove that the numbers in the right-hand side of (1.1) are exactly the moments of the standard Gaussian measure γ:

$$\int_{-\infty}^{\infty} \lambda^k e^{-\lambda^2/2} \frac{d\lambda}{\sqrt{2\pi}} = \begin{cases} \frac{k!}{2^{k/2}(k/2)!} \ , & k \text{ is even} \\ 0 \ , & k \text{ is odd} \end{cases}$$

Therefore, it is natural to ask whether (1.1) implies the weak convergence of μ_n to the Gaussian measure γ, i.e.

$$\forall \phi \in C_{\mathrm{bdd}}(\mathbb{R}) \qquad \lim_{n\to\infty} \int \phi(\lambda) d\mu_n(\lambda) \overset{??}{=} \int \phi(\lambda) e^{-\lambda^2/2} \frac{d\lambda}{\sqrt{2\pi}} .$$

This was solved in the affirmative by Andrei Markov, who developed Chebyshev's ideas.

[1] In Chebyshev's formulation, explicit bounds for finite n played a central rôle. Markov's form of Theorem 1.4 was also stated more quantitatively than here.

Theorem 1.4 (Markov). *Let $(\mu_n)_{n\geq 1}$ be a sequence of probability measures that satisfy (1.1). Then $\mu_n \underset{n\to\infty}{\longrightarrow} \gamma$ in weak topology.*

In juxtaposition with the two exercises above, this theorem implies the Central Limit Theorem for random variables with finite moments[2].

The theorem follows from two propositions. The first one shows that the crucial property of the limiting measure is determinacy.[3]

Proposition 1.5 (Fréchet–Shohat). *If μ is a determinate measure (i.e. its moments determine it uniquely), and the measures μ_n are such that*

$$\forall k \geq 0 \quad \lim_{n\to\infty} s_k[\mu_n] = s_k[\mu] \ , \tag{1.2}$$

then $\mu_n \to \mu$ in weak topology.

Obviously[4], if μ is indeterminate, then one can find $\mu_n \not\to \mu$ such that (1.2) holds.

How to check whether a measure is determinate?

Exercise 1.6. Prove that every compactly supported measure is determinate.

To handle the Gaussian measure, we need to relax the assumptions.

Proposition 1.7. *If μ is such that*

$$\exists \epsilon > 0 : \quad \int e^{\epsilon|\lambda|} d\mu(\lambda) < \infty \ , \tag{1.3}$$

then μ is determinate.

Obviously[5], Proposition 1.5 and Proposition 1.7 imply Theorem 1.4.

Exercise 1.8. The condition (1.3) (called: "μ has exponential tails") is equivalent to the following condition on the moments ("factorial growth"):

$$\exists C > 0 : \quad \forall k \geq 0 \quad s_{2k}[\mu] \leq C^{k+1}(2k)! \ . \tag{1.4}$$

The exponential tails are not a necessary condition for determinacy. In fact, there exist determinate measures with extremely heavy tails.[6] However, it is important to know that not all measures are determinate.

[2] The Lévy–Lindeberg CLT can be also obtained by the combination of such arguments with a truncation.

[3] It is the proof of this proposition that relies on a compactness argument which Chebyshev may not have approved.

[4] i.e. please convince yourself that this is obvious before proceeding

[5] in the sense of footnote 4

[6] i.e. determinacy does not imply anything stronger than the obvious condition $\mu(\mathbb{R}\setminus[-R,R]) = O(R^{-\infty})$

Example 1.9. *Let $u(y)$ be a 1-periodic function (say, a bounded and measurable one). Denote*

$$Z_u = \int \exp(-\frac{1}{2}y^2 + u(y))dy$$

and define a measure supported on $\mathbb{R}_+$:

$$d\mu_u(\lambda) = \frac{1}{Z_u} \exp(-\frac{1}{2}\log^2\lambda + u(\log\lambda))\frac{d\lambda}{\lambda} .$$

Then $s_k[\mu_n] = e^{k^2/2}$ for all k, regardless of the choice of u! In particular, none of these measures is determinate.

Exercise 1.10. (a) Prove this. (b) Is there a discrete measure with the same moments as these μ_u?

1.2 Proofs of Proposition 1.5 and 1.7

Proof of Proposition 1.5. It suffices to show that (a) $(\mu_n)_{n\geq 1}$ is precompact in weak topology, and (b) μ is the unique weak limit point of this sequence. To prove (a), recall the criterion for compactness in weak topology (in the old days, it used to be called the Helly selection theorem): a collection $\mathfrak{M}$ of finite measures is precompact if and only if the following two conditions hold:

$$\sup_{\nu\in\mathfrak{M}} \nu(\mathbb{R}) < \infty \tag{1.5}$$

$$\forall\epsilon > 0 \ \exists R > 0 : \ \sup_{\nu\in\mathfrak{M}} \nu(\mathbb{R}\setminus[-R,R]) < \epsilon \tag{1.6}$$

The first condition holds for $\mathfrak{M} = \{\mu_n\}$ since

$$\lim_{n\to\infty} \mu_n(\mathbb{R}) = \lim_{n\to\infty} s_0[\mu_n] = s_0[\mu] < \infty ,$$

whereas the second one follows from the Chebyshev inequality:

$$\mu_n(\mathbb{R}\setminus[-R,R]) \leq \frac{\sup_n s_2[\mu_n]}{R^2} \leq \frac{C}{R^2} .$$

Thus (a) is proved and we proceed to (b). If ν is a limit point of (μ_n), we have:

$$\mu_{n_j} \to \nu , \quad j\to\infty .$$

By definition

$$\int \phi(\lambda)d\mu_{n_j}(\lambda) \to \int \phi(\lambda)d\nu(\lambda)$$

for any bounded continuous ϕ. Please check (using the assumptions) that also

$$\int \lambda^k d\mu_{n_j}(\lambda) \to \int \lambda^k d\nu(\lambda) ,$$

although the function $\lambda \mapsto \lambda^k$ is not bounded for $k \geq 1$. Then we have:

$$s_k[\mu] = \lim_{n\to\infty} s_k[\mu_n] = s_k[\nu]$$

for all k, and therefore $\nu = \mu$ by determinacy. $\square$

The proof of Proposition 1.7 is a bit more analytic.

Proof of Proposition 1.7. Consider the Fourier–Stieltjes transform of μ, which is[7] the function

$$\phi(\xi) = \int e^{i\xi\lambda} d\mu(\lambda) \ .$$

The integral converges for complex ξ with $|\Im\xi| < \epsilon$ (where ϵ comes from the definition (1.3) of exponential tails), therefore ϕ can be extended to an analytic function in this strip. If ν is another measure with the same moments, then by Exercise 1.8 also

$$\psi(\xi) = \int e^{i\xi\lambda} d\nu(\lambda)$$

is analytic in a strip.[8] Now observe that

$$\phi^{(k)}(0) = i^k s_k[\mu] = \psi^{(k)}(0) \ , \quad k = 0, 1, 2, \cdots$$

and hence, by the uniqueness theorem for analytic functions, $\psi \equiv \phi$ in some strip containing $\mathbb{R}$ and in particular on $\mathbb{R}$. Invoking the inversion formula for the Fourier–Stieltjes transform, we infer that $\nu = \mu$. $\square$

2 Quasianalyticity; determinacy

While the condition of exponential tails in Proposition 1.7 can not be dropped, it can be relaxed. The proof relied on a uniqueness theorem for analytic functions, so we shall discuss wider classes of functions (known as quasianalytic classes) in which one has a uniqueness theorem. The idea goes back to Hadamard, and the theory was developed in the first half of the XX-th century. Two of the classical references are Carleman [1926], Mandelbrojt [1942].

2.1 The Denjoy–Carleman theorem

First, let us move back to the real domain. How do we state the uniqueness theorem for analytic functions without using complex variables?

---

[7]The properties of the Fourier–Stieltjes transform are described in any textbook on harmonic analysis, e.g. Katznelson [2004]. We shall only use the fact that a finite measure is determined by its transform, as follows from the inversion formula.

[8]formally, this may be a smaller strip $|\Im\xi| < \epsilon' \leq \epsilon$; this is sufficient for our purposes, though in fact one may take $\epsilon' = \epsilon$.

Exercise 2.1. Let $I \subset \mathbb{R}$ be a bounded interval. A function $\phi \in C^\infty(I)$ has an analytic extension to a neighbourhood of I if and only if there exists $C > 0$ such that for every $k \geq 0$ and $\xi \in I$

$$|\phi^{(k)}(\xi)| \leq C^{k+1} k! \tag{2.1}$$

Corollary 2.2. *Let $C > 0$, and let $\phi, \psi \in C^\infty(\mathbb{R})$ be two functions satisfying (2.1). If $\phi^{(k)}(0) = \psi^{(k)}(0)$ for all $k \geq 0$, then $\phi \equiv \psi$.*

Exercise 2.3. Prove the corollary without using the theory of functions of a complex variable.

In the proof of Proposition 1.7, we could have used the uniqueness theorem in this form. We know from a calculus course that the assumption (2.1) can not be dropped: there exist non-zero C^∞ functions which vanish at a point with all derivatives. However, it can be relaxed.

Definition 2.4. *Let $\mathcal{M} = (M_k)_{k \geq 0}$ be a sequence of positive numbers. The Carleman class $C\{\mathcal{M}\}$ consists of all $\phi \in C^\infty(\mathbb{R})$ such that, for some $C > 0$,*

$$\sup_{\xi \in \mathbb{R}} |\phi^{(k)}(\xi)| \leq C^{k+1} M_k \ . \tag{2.2}$$

Definition 2.5. *A Carleman class is called quasianalytic if the map*

$$\phi \mapsto (\phi^{(k)}(0))_{k \geq 0}$$

is injective, i.e. if a function vanishing with all derivatives at a point has to vanish identically.

Example 2.6. *For $M_k = k!$, the class $C\{\mathcal{M}\}$ is quasianalytic.*

Exercise 2.7. Find an explicit non-quasianalytic Carleman class (without reverting to the Denjoy–Carleman theorem below).

From now on we shall assume that the sequence $\mathcal{M}$ is log-convex:

$$M_k \leq \sqrt{M_{k-1} M_{k+1}} \ . \tag{2.3}$$

This regularity assumption does not entail a great loss of generality: every Carleman class $C\{\mathcal{M}\}$ can be embedded in a larger (explicit) Carleman class $C\{\mathcal{M}'\}$ which satisfies (2.3), so that $\mathcal{C}\{\mathcal{M}\}$ is quasianalytic if and only if $C\{\mathcal{M}'\}$ is quasianalytic (see Mandelbrojt [1942]).

Theorem 2.8 (Denjoy–Carleman)**.** *Let $\mathcal{M}$ be a log-convex (2.3) sequence of positive numbers. The Carleman class $C\{\mathcal{M}\}$ is quasianalytic if and only if*

$$\sum_{k \geq 0} \frac{M_k}{M_{k+1}} = \infty \ . \tag{2.4}$$

Example 2.9. *Each of the following sequences defines a quasianalytic Carleman class $C\{\mathcal{M}\}$:*

$$M_k = k!\log^k(k+10)\ ,\ M_k = k!\log^k(k+10)\log^k\log(k+10)\ ,\cdots$$

On the other hand, $M_k = k!\log^{2k}(k+10)$ defines a non-quasianalytic class.

Remark 2.10. Let $\mathcal{M}$ be a sequence of positive numbers. If

$$\sum_{k\geq 1} M_k^{-1/k} = \infty\ , \tag{2.5}$$

then (2.4) holds. Vice versa, if $\mathcal{M}$ satisfies (2.4) and (2.3), then (2.5) holds.

Proof of Remark 2.10. Assume that $\mathcal{M}$ is log-convex. Then $M_k \leq M_0(M_k/M_{k-1})^k$, hence

$$\sum_{k\geq 1} M_k^{-1/k} \geq \sum_{k\geq 1} M_0^{-1/k}\frac{M_{k-1}}{M_k} \geq \min(\frac{1}{M_0},1)\sum_{k\geq 1}\frac{M_{k-1}}{M_k}\ .$$

The condition (2.4) implies that the right-hand side is infinite, and hence (2.5) holds. The reverse implication (2.5)$\Longrightarrow$(2.4) follows from

Lemma 2.11 (Carleman's inequality)**.** *For any positive sequence (a_k),*

$$\sum_{k=1}^{\infty}(a_1\cdots a_k)^{1/k} \leq e\sum_{k=1}^{\infty} a_k\ . \tag{2.6}$$

Proof. (See [Pólya, 1990, Chapter XVI] for a discussion) Let r_k be auxiliary positive numbers. Then

$$(a_1\cdots a_k)^{1/k} = \left(\frac{a_1r_1\cdots a_kr_k}{r_1\cdots r_k}\right)^{1/k} \leq \frac{a_1r_1+\cdots+a_kr_k}{k(r_1\cdots r_k)^{1/k}}\ .$$

Therefore

$$\sum_{k=1}^{\infty}(a_1\cdots a_k)^{1/k} \leq \sum_{k=1}^{\infty}\frac{a_1r_1+\cdots+a_kr_k}{k(r_1\cdots r_k)^{1/k}} = \sum_{j=1}^{\infty}a_jr_j\sum_{k\geq j}\frac{1}{k(r_1\cdots r_k)^{1/k}}\ .$$

Choose $r_k = (k+1)^k/k^{k-1}$ so that $r_1r_2\cdots r_k = (k+1)^k$; then the right-hand side takes the value

$$\begin{aligned}\sum_{j=1}^{\infty}a_j\frac{(j+1)^j}{j^{j-1}}\sum_{k\geq j}\frac{1}{k(k+1)} &= \sum_{j=1}^{\infty}a_j\frac{(j+1)^j}{j^{j-1}}\frac{1}{j}\\ &= \sum_{j=1}^{\infty}a_j(1+1/j)^j \leq e\sum_{j=1}^{\infty}a_j\ .\end{aligned}$$

This concludes the proof of the Carleman inequality (2.6) and of the remark.

□

Corollary 2.12 (Carleman's criterion)**.** *Any measure* μ *with*

$$\sum_{k=0}^{\infty} s_{2k}[\mu]^{-\frac{1}{2k}} = \infty \ . \tag{2.7}$$

is determinate.

Proof of Carleman's criterion. Let $\mathcal{M} = (M_k)_{k\geq 0}$, where $M_k = \sqrt{s_{2k}[\mu] s_0[\mu]}$. By Hölder's inequality, $\mathcal{M}$ is log-convex (2.3). Let

$$\phi(\xi) = \int e^{i\xi\lambda} d\mu(\lambda) \ ,$$

then

$$\sup_{\xi} |\phi^{(k)}(\xi)| \leq \int |\lambda|^k d\mu(\lambda) \leq M_k \ .$$

The same bound is satisfied by the Fourier transform of any measure ν sharing the moments of μ. By the Denjoy–Carleman theorem (and Remark 2.10) the class $C\{\mathcal{M}\}$ is quasianalytic, hence ϕ (and thus also μ) is uniquely determined by the moments s_k. □

Now we prove the Denjoy–Carleman theorem. Before proceeding to the proof, the readers may wish to to convince themselves that the uniqueness part (sufficiency) is not a direct consequence of the Taylor expansion with remainder. In fact, the original proofs relied on complex-variable methods. The first two real-variable proofs were found by Bang [1946, 1953]; we reproduce the second one.[9]

Proof of Theorem 2.8.

Sufficiency First let us prove that (2.4) implies quasianalyticity, following an argument due to Bang [1953].

Assume that $\phi \in C^\infty$ admits the bounds

$$\sup_{\xi\in\mathbb{R}} |\phi^{(k)}(\xi)| \leq C^{k+1} M_k \tag{2.8}$$

For integer $p \geq 0$, denote

$$\mathcal{B}_p = \left\{ \xi \in \mathbb{R} \,\middle|\, \forall 0 \leq k < p \ \ |\phi^{(k)}(\xi)| \leq C^{k+1} e^{k-p} M_k \right\} \ . \tag{2.9}$$

Note that $\mathbb{R} = \mathcal{B}_0 \supset \mathcal{B}_1 \supset \mathcal{B}_2 \cdots$. If ϕ vanishes with all the derivatives at 0, then $0 \in \bigcap_{p\geq 0} \mathcal{B}_p$. The following lemma asserts that if a point lies in $\mathcal{B}_p$, then its neighbourhood lies in $\mathcal{B}_{p-1}$. (Properties of this kind are sometimes called "propagation of smallness".) It will be more convenient to state it in the contrapositive:

[9]Thøger Sophus Vilhelm Bang (27.6.1917–18.1.1997) — a Danish mathematician. In addition to his work on quasianalyticity, Bang is remembered for the solution of Tarski's plank problem.

Lemma 2.13. *Assume that ϕ satisfies (2.8) for some log-convex sequence $\mathcal{M}$. If $\xi \notin \mathcal{B}_{p-1}$ for some $p \geq 1$, then*

$$\left[\xi - \frac{M_{p-1}}{CeM_p}, \xi + \frac{M_{p-1}}{CeM_p}\right] \cap \mathcal{B}_p = \varnothing \ .$$

Proof of Lemma 2.13. Assume that $\xi + h \in \mathcal{B}_p$ for some $|h| \leq \frac{M_{p-1}}{CeM_p}$. Then for $k \leq p-1$

$$|\phi^{(k)}(\xi)| \leq \sum_{j=0}^{p-k-1} |\phi^{(k+j)}(\xi+h)| \frac{|h|^j}{j!} + C^{p+1} |\phi^{(p)}(\xi + \tilde{h})| \frac{|h|^{p-k}}{(p-k)!} \tag{2.10}$$

$$\leq \sum_{j=0}^{p-k} C^{k+j+1} e^{k+j-p} M_{k+j} \frac{|h|^j}{j!} \ , \tag{2.11}$$

where on the second step we used (2.9) to bound the terms in the sum and (2.8) to bound the remainder. Using log-convexity, we bound

$$\begin{aligned}(2.11) &= M_k \sum_{j=0}^{p-k} C^{k+j+1} e^{k+j-p} \frac{M_{k+j}}{M_k} \frac{|h|^j}{j!} \\ &\leq M_k \sum_{j=0}^{p-k} C^{k+j+1} e^{k+j-p} \left(\frac{M_p}{M_{p-1}}\right)^j \frac{|h|^j}{j!} \leq M_k \sum_{j=0}^{\infty} [\cdots] \\ &= M_k C^{k+1} e^{k-p} \exp\left\{ Ce|h| \frac{M_p}{M_{p-1}} \right\} \leq M_k C^{k+1} e^{k-(p-1)}\end{aligned}$$

This proves the lemma. □

Now we conclude the proof of sufficiency. If ϕ is not identically zero, there exist p and ξ such that $\xi \notin \mathcal{B}_p$. By Lemma 2.13, for $q > p$

$$\left[\xi - \sum_{k=p+1}^{q} \frac{M_{k-1}}{CeM_k}, \xi + \sum_{k=p+1}^{q} \frac{M_{k-1}}{CeM_k}\right] \cap \mathcal{B}_{p+k} = \varnothing \ ,$$

thus – by the main assumption (2.4) – we have $0 \notin \mathcal{B}_q$ for sufficiently large q. Therefore ϕ can not vanish at zero with all the derivatives. This proves the sufficiency of the Denjoy–Carleman condition (2.4) for quasianalyticity.

Necessity Now we assume that (2.4) fails and construct a non-zero compactly supported $\phi \in C\{\mathcal{M}\}$. The construction goes back at least to Mandelbrojt [1942], where unpublished work of Bray is quoted.

Let u be a bump function such that

$$0 \leq u \leq 1 \ , \quad \operatorname{supp} u \subset [-2, 2] \ , \quad \int u(\xi) d\xi = 1 \ , \quad \int |u'(\xi)| d\xi \leq 1$$

In fact, $u = \frac{1}{2}\mathbb{1}_{[-1,1]}$ is fine, but you may as well take a C^1 function if δ-functions make you feel uncomfortable. Let $M_{-1} = M_0^2/M_1$; define $u_k(\xi) = \frac{M_k}{M_{k-1}} u(\xi \frac{M_k}{M_{k-1}})$, and let

$$\phi_p = u_0 * u_1 * \cdots * u_{p-1}$$

be the convolution of the first p functions. Then, for $p > k$, ϕ_p admits the bounds:

$$\begin{aligned} |\phi_p^{(k)}| &= |u_0' * u_1' * \cdots * u_{k-1}' * u_k * \cdots * u_{p-1}| \\ &\leq \prod_{j=0}^{k-1} \int |u_j'(\xi)| d\xi \times \max_\xi |u_k(\xi)| \times \prod_{j=k+1}^{p-1} \int u_j(\xi) d\xi \\ &\leq \prod_{j=0}^{k-1} \frac{M_j}{M_{j-1}} \times \frac{M_k}{M_{k-1}} = \frac{M_k}{M_{-1}} \end{aligned} \tag{2.12}$$

In particular, for any k the sequence $(\phi_p^{(k)})_{p>k}$ is precompact in uniform topology. Choose ϕ such that for any k the derivative $\phi^{(k)}$ is a uniform limit point of $(\phi_p^{(k)})_{p>k}$.[10] The estimates (2.12) allow to exchange the limit with differentiation, therefore

$$\forall k \geq 0 \quad |\phi^{(k)}| \leq \frac{M_k}{M_{-1}} \ ;$$

also,

$$\operatorname{supp} \phi \subset \left[-2 \sum_{k \geq 0} \frac{M_{k-1}}{M_k}, 2 \sum_{k \geq 0} \frac{M_{k-1}}{M_k} \right] \subsetneq \mathbb{R} \ .$$

□

Remark 2.14. The proof of necessity shows that if (2.4) fails, one may construct non-zero functions in the corresponding Carleman class with a prescribed constant C in (2.2). Alternatively, one may prescribe the support of the function. Also, the log-convexity assumption (2.3) is not used in the proof of necessity.

2.2 Remarks

Sharpness Although Carleman's condition (2.7) is not necessary for the determinacy of the moment problem, it is sharp in the following sense.[11]

Proposition 2.15. *If μ is a measure that fails (2.7), then there exists an indeterminate measure ν such that $s_{2k}[\nu] \leq s_{2k}[\mu]$ for all $k \geq 0$.*

Proof. Suppose μ fails (2.7). Then the class of functions ϕ with

$$|\phi^{(2n)}|, |\phi^{(2n+1)}| \leq s_{2n}[\mu]/(2\pi)^2$$

is not quasianalytic[12], and thus contains a non-zero function ϕ with $\operatorname{supp} \phi \subset [1, A]$ for

[10] It is not hard to see that ϕ is unique, so in fact $\phi_p^{(k)} \to \phi^{(k)}$ for any k, but we do not need this.

[11] Cf. [Kostyučenko and Mityagin, 1960, Theorem 6]. I learned the argument below from B. Mityagin, who refused to take credit for it: ≪ "everybody" – say, Sz. Mandelbrojt or B. Ya. Levin – knew this in the 30's, maybe without saying it explicitly all the time.≫

[12] Why? note that the condition (2.3) may fail, and find a way to save the argument

some $A > 1$ (cf. Remark 2.14). Decompose the Fourier transform $\hat{\phi}$ into a difference $\hat{\phi} = g - h$ of two non-negative functions, then the measures with Radon densities $g/\sqrt{A}$ and $h/\sqrt{A}$ have the same moments, and these moments can be bounded as follows. First, by Cauchy–Schwarz

$$\begin{aligned} \int \lambda^{2n} g(\lambda) d\lambda \leq \int \lambda^{2n} |\hat{\phi}(\lambda)| d\lambda = \int \lambda^{2n} \sqrt{1+\lambda^2} |\hat{\phi}(\lambda)| \frac{d\lambda}{\sqrt{1+\lambda^2}} \\ \leq \left\{ \pi \int \lambda^{4n} (1+\lambda^2) |\hat{\phi}(\lambda)|^2 d\lambda \right\}^{1/2} . \end{aligned} \tag{2.13}$$

Using the Parseval identity,

$$(2.13) \leq \left\{ 2\pi^2 \int \left[|\phi^{(2n)}(\lambda)|^2 + |\phi^{(2n+1)}(\lambda)|^2 \right] d\lambda \right\}^{1/2} \leq \sqrt{A \, s_{2n}[\mu]} \ . \tag{2.14}$$

Thus the moments of the indeterminate measure ν, $d\nu(\lambda) = A^{-1/2} g(\lambda) d\lambda$, are majorised by those of μ. □

3 Existence

The Hamburger moment problem asks whether there exists a (positive Borel) measure μ on $\mathbb{R}$ with the given sequence of moments $(s_k)_{k \geq 0}$. There is an obvious necessary condition: if such μ exists, one has for any k and any $z_0, \cdots, z_k \in \mathbb{C}$:

$$\sum_{j=0}^{k} \sum_{l=0}^{k} s_{j+l} z_j \bar{z}_l = \int |\sum_{j=0}^{k} z_j \lambda^j|^2 d\mu(\lambda) \geq 0 \ .$$

In other words, the Hankel matrix $H = (s_{j+l})_{j,l=0}^{\infty}$ should be positive semidefinite, i.e. define a positive semidefinite quadratic form.

3.1 Hamburger's theorem

Theorem 3.1 (Hamburger). *A sequence $(s_k)_{k \geq 0}$ is a moment sequence if and only if the corresponding Hankel matrix $H = (s_{j+l})_{j,l=0}^{\infty}$ is positive semidefinite, i.e. for any k and any $z_0, z_1, \cdots, z_k \in \mathbb{C}$*

$$\sum_{j,l=0}^{k} s_{j+l} z_j \bar{z}_l \geq 0 \ .$$

In the sequel, we employ the following (ab)use of notation. If $R(\lambda) = \sum c_j \lambda^j$ is a polynomial, we denote $\overline{R}(\lambda) = \sum \overline{c_j} \lambda^j$ and $|R|^2 = R\bar{R}$. The first step in the proof is

Proposition 3.2. *If $P \in \mathbb{C}[\lambda]$ is non-negative on $\mathbb{R}$, then it can be represented as $P = |R|^2$ for some $R \in \mathbb{C}[\lambda]$.*

Proof. From the assumption P has real coefficients. Let a_j be the real zeros of P, so that a zero of multiplicity $2m$ is counted m times, and let b_l be the complex zeros of P in the upper half-plane. Then (for real λ)

$$P(\lambda) = c \prod_j (\lambda - a_j)^2 \prod_l |\lambda - b_l|^2$$

for some $c > 0$. Then $P = |R|^2$ for

$$R(\lambda) = \sqrt{c} \prod_j (\lambda - a_j) \prod_l (\lambda - b_l) \ . \qquad \square$$

To prove Hamburger's theorem, define a functional $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ by

$$\Phi[\sum_j a_j \lambda^j] = \sum_j a_j s_j \ .$$

It is positive in the following sense: $\Phi[P] \geq 0$ whenever P is a non-negative polynomial. Let E be the linear space spanned by $\mathbb{C}[\lambda]$ and the collection of functions $\{\mathbb{1}_{(-\infty,\lambda]}\}_{\lambda\in\mathbb{R}}$, and let $K \subset E$ be the cone of non-negative functions. We shall extend Φ to a linear functional $\widetilde{\Phi} : E \to \mathbb{C}$ such that $\Phi(K) \subset \mathbb{R}_+$. This is done using the following general device.

Theorem 3.3 (M. Riesz)**.** *Let E be a linear space and let $K \subset E$ be a convex cone. If $\Phi : F \to \mathbb{C}$ is a linear functional defined on a linear subspace $F \subset E$ so that $\Phi(F \cap K) \subset \mathbb{R}_+$ and $E = F + K$ (i.e. every element of E can be represented as a sum of an element of F and an element of K), then there exists a linear extension $\widetilde{\Phi} : E \to \mathbb{C}$, $\widetilde{\Phi}|_F = \Phi$, such that $\widetilde{\Phi}(K) \subset \mathbb{R}_+$.*

Proof of the Hamburger theorem. Let us apply the Riesz theorem to

$$F = \mathbb{C}[\lambda] \ , \ \ E = \mathbb{C}[\lambda] + \mathrm{span}\{\mathbb{1}_{(-\infty,\lambda]}\}_{\lambda\in\mathbb{R}} \ , \ \ K = \{\phi \in E \,|\, \phi(\mathbb{R}) \subset \mathbb{R}_+\}$$

and $\Phi[\sum_j a_j \lambda^j] = \sum_j a_j s_j$ as above.

Exercise 3.4. Check that $E = F + K$.

By the Riesz extension theorem, there exists an extension $\widetilde{\Phi} : E \to \mathbb{C}$ such that $\widetilde{\Phi}(K) \subset \mathbb{R}_+$. Denote $\mu(\lambda) = \widetilde{\Phi}(\mathbb{1}_{(-\infty,\lambda]})$; then $\mu : \mathbb{R} \to \mathbb{R}$ is non-negative, non-decreasing and bounded. Let us show that the corresponding measure (of which μ is the cumulative distribution function) is the desired solution of the moment problem, i.e. that

$$\forall P \in \mathbb{C}[\lambda] \qquad \int P(\lambda) d\mu(\lambda) = \Phi[P] \ . \qquad (3.1)$$

Representing

$$P = \frac{P^2 + 1}{2} - \frac{(P-1)^2}{2} \ ,$$

we reduce (3.1) to its special case $P \geq 0$. Choose a sequence (f_j) in

$$\mathrm{span}\{\mathbb{1}_{(-\infty,a]} \mid \mu\{a\} = 0\}$$

such that

$$0 \leq f_j \leq P \ , \quad f_j \nearrow P \quad \text{locally uniformly.}$$

Then by monotone convergence

$$\int P d\mu = \lim_{j\to\infty} \int f_j d\mu = \lim_{j\to\infty} \widetilde{\Phi}[f_j] \leq \lim_{j\to\infty} \widetilde{\Phi}[P] = \Phi[P] \ .$$

In the converse direction, choose $\epsilon > 0$ and choose $R > 0$ such that $P(\lambda) \geq \frac{1}{\epsilon}$ for $|\lambda| \geq R$. Then, for sufficiently large j, $f_j \geq P - \epsilon$ on $[-R, R]$, whereas for $|\lambda| \geq R$ we have $P(\lambda) \leq \epsilon P(\lambda)^2$. Thus for all λ

$$P(\lambda) \leq f_j(\lambda) + \epsilon + \epsilon P(\lambda)^2 \ ,$$

and therefore

$$\begin{aligned}\Phi[P] &\leq \widetilde{\Phi}[f_j] + \epsilon\Phi[1] + \epsilon\Phi[P^2] \\ &= \int f_j d\mu + \epsilon(\Phi[1] + \Phi[P^2]) \leq \int P d\mu + \epsilon(\Phi[1] + \Phi[P^2]) \ .\end{aligned}$$

Letting $\epsilon \to +0$, we obtain that $\Phi[P] \leq \int P d\mu$. □

Proof of the Riesz theorem. Let us first consider the case when $\dim E/F = 1$. Let $E = F + u\mathbb{C}$. To define $\widetilde{\Phi}(u) = a$, we need to satisfy the constraints

$$\inf_{f\in F, f+u\in K} \Phi(f) + a \geq 0 \ , \qquad \inf_{f\in F, f-u\in K} \Phi(f) - a \geq 0 \ ,$$

which are equivalent to

$$-\inf_{f_1\in F, f_1+u\in K} \Phi(f_1) \leq a \leq \inf_{f_2\in F, f_2-u\in K} \Phi(f_2) \ . \tag{3.2}$$

According to the assumption $F + K = E$, both infima are taken over non-empty sets. Furthermore, for each such f_1 and f_2

$$\Phi(f_2) + \Phi(f_1) = \Phi(f_2 - u) + \Phi(f_1 + u) \geq 0 \ ,$$

hence the right-hand side of (3.2) is not smaller than the left-hand side, hence a good choice of a exists.

The general case follows by transfinite induction. Consider the set of pairs $\mathfrak{P} = \{(F', \Phi')\}$, where $F \subset F' \subset E$ is a linear space, and $\Phi' : F' \to \mathbb{C}$ is a linear functional such that $\Phi'|_F = \Phi$ and $\Phi'(K \cap F') \subset \mathbb{R}_+$. Introduce a partial order:

$$(F', \Phi') \prec (F'', \Phi'') \quad \text{if} \quad F' \subset F \wedge \Phi''|_{F'} = \Phi' \ .$$

If $\mathfrak{C} \subset \mathfrak{P}$ is a chain (linearly ordered subset), then

$$\left(F'' = \bigcup_{(F',\Phi')\in\mathfrak{C}} F', \ \Phi'' = \bigcup_{(F',\Phi')\in\mathfrak{C}} \Phi' \right) \in \mathfrak{C} \ ;$$

hence by Zorn's lemma $\mathfrak{P}$ has a maximal element $(\widetilde{F}, \widetilde{\Phi})$. By the special case considered above, $\widetilde{F} = E$, and the theorem is proved. □

Exercise 3.5. Does the conclusion of the theorem hold without the assumption $E = F + K$? Prove or construct a counterexample.

Exercise 3.6. Show that the Riesz extension theorem implies the Hahn–Banach theorem: if $F \subset E$ are linear spaces, $\| \cdot \|$ is a seminorm on E and $\Psi : F \to \mathbb{C}$ satisfies $|\Psi| \leq \| \cdot \|$, then there is a linear extension $\widetilde{\Psi} : E \to \mathbb{C}$ of Ψ such that $|\widetilde{\Psi}| \leq \| \cdot \|$.

Remark 3.7. By Sylvester's criterion, strictly positive definite Hankel matrices are characterised by the sequence of inequalities

$$\det(s_{j+l})_{j,l=0}^k > 0 \ , \quad k \geq 0 \ .$$

These matrices correspond to measures the support of which contains an infinite number of points.

Remark 3.8. The corresponding characterisation of moment sequences corresponding to measures supported on a finite set of points was found by Berg and Szwarc [2015]: there exists k_0 such that

$$\det(s_{j+l})_{j,l=0}^k \begin{cases} > 0, & k < k_0 \\ = 0, & k \geq k_0 \end{cases}$$

3.2 The spectral theorem for bounded self-adjoint operators

Let $\mathcal{H}$ be a Hilbert space, and let $T : \mathcal{H} \to \mathcal{H}$ be a bounded self-adjoint operator. This means that $\|Tu\| \leq K\|u\|$ for any $u \in \mathcal{H}$ (the smallest $K \geq 0$ for which this inequality holds is the operator norm $\|T\|$), and

$$\langle Tu, v \rangle = \langle u, Tv \rangle$$

for any $u, v \in \mathcal{H}$.

Theorem 3.9 (The spectral theorem)**.** *There exist a collection $\{\mu_\alpha\}_{\alpha\in\mathcal{A}}$ of Borel probability measures on the real line and a unitary (i.e. norm-preserving) bijection*

$$U : \mathcal{H} \longleftrightarrow \bigoplus_{\alpha\in\mathcal{A}} L_2(\mu_\alpha)$$

which conjugates T to a direct sum of multiplication operators:

$$UTU^{-1} : \bigoplus_{\alpha\in\mathcal{A}} L_2(\mu_\alpha) \to \bigoplus_{\alpha\in\mathcal{A}} L_2(\mu_\alpha) \ , \quad (f_\alpha(\lambda))_{\alpha\in\mathcal{A}} \mapsto (\lambda f_\alpha(\lambda))_{\alpha\in\mathcal{A}} \ . \qquad (3.3)$$

The measures μ_α are supported in $[-\|T\|, \|T\|]$, and in fact

$$\bigcup_\alpha \operatorname{supp} \mu_\alpha = \sigma(T) \quad \textit{(the spectrum of } T\textit{)} \tag{3.4}$$

The following exercise may help digest the formulation of the theorem, if you have not seen it before.

Exercise 3.10. (a) Check that the theorem holds in $\dim \mathcal{H} < \infty$. (b) $\sum_\alpha \mu_\alpha$ is pure point if and only if T has an orthonormal basis of eigenvectors.

Now we proceed to the proof of the theorem.

Lemma 3.11. *Let $u \in \mathcal{H}$. There exists a measure μ_u supported in $[-\|T\|, \|T\|]$ such that $\langle T^k u, u\rangle = s_k[\mu_u]$ for any $k \geq 0$.*

Proof. Observe that

$$\sum_{j,\ell} \langle T_{j+\ell} u, u\rangle z_j \bar{z}_\ell = \| \sum_j z_j T^j u\|^2 \geq 0$$

and use the Hamburger theorem to construct μ. Then

$$\int \lambda^{2k} d\mu(\lambda) \leq \|T\|^{2k} \ ,$$

hence $\operatorname{supp} \mu \subset [-\|T\|, \|T\|]$ (why?) □

Let us now explain why the lemma implies the spectral theorem. For $u \in \mathcal{H}$, define

$$U_u : \operatorname{span} \left\{ T^k u \mid k \geq 0 \right\} \to L_2(\mu_u)$$

which sends $T^k u$ to the monomial $\lambda^k \in L_2(\mu_u)$. Let us check that this operator is norm-preservng:

$$\begin{aligned}
\|(\sum a_k T^k u)\|^2 &= \sum_{k,\ell} a_k \bar{a}_l \langle T^k u \ , \ T^l u\rangle \\
&= \sum_{k,\ell} a_k \bar{a}_l \langle T^{k+l} u \ , \ u\rangle && \text{(by self-adjointness)} \\
&= \sum_{k,\ell} a_k \bar{a}_l s_{k+l}[\mu] && \text{(by construction)},
\end{aligned}$$

which is equal to

$$\|U(\sum a_k T^k u)\|^2 = \sum_{k,\ell} a_k \bar{a}_l \int \lambda^k \lambda^l d\mu_u(\lambda) \ .$$

In particular U_u can be extended to the closure $\mathcal{H}_u = \overline{\operatorname{span} \{T^k u \mid k \geq 0\}}$. Then $U_u : \mathcal{H}_u \leftrightarrow L_2(\mu_u)$ is a unitary bijection (why is it onto?), and

$$(U_u T|_{\mathcal{H}_u} U_u^{-1} f)(\lambda) = \lambda f(\lambda)$$

since this equality holds for $f(\lambda) = \lambda^k$. This proves the theorem in the case $\mathcal{H}_u = \mathcal{H}$. The equality (3.4) follows a posteriori from (3.3) and the fact that the spectrum of a multiplication operator $f(\lambda) \mapsto \lambda f(\lambda)$ in $L_2(\mu)$ is exactly the support of μ.

Exercise 3.12. Complete the proof of the spectral theorem in full generality.

Exercise 3.13. Compute the spectral measure μ_u for

$$T : \ell_2(\mathbb{Z}) \to \ell_2(\mathbb{Z}) \ , \quad (a_k)_{k\in\mathbb{Z}} \mapsto \left(\sum_{j=0}^{m} [\alpha_j a_{k+j} + \bar{\alpha}_j a_{k-j}] \right)_{k\in\mathbb{Z}}$$

where $(\alpha_j)_{j=0}^m$ are fixed complex coefficients, and

$$u = (u_j)_{j\in\mathbb{Z}} \ , \quad u_j = \begin{cases} 1 \ , & j = 0 \\ 0 \ . \end{cases}$$

3.3 Other moment problems*

The Stieltjes moment problem The Stieltjes moment problem concerns the moment sequences of measures that are supported on a half-line. Chronologically, the following result preceded Hamburger's theorem. The original argument of Stieltjes was based on considerations involving continued fractions, which can not be directly applied to the moment problem on the line.

Theorem 3.14 (Stieltjes)**.** *A sequence (s_k) is the moment sequence of a measure μ supported on $\mathbb{R}_+$ if and only if*

$$H = (s_{j+l})_{j,l=0}^{\infty} \succeq 0 \quad \text{and} \quad H' = (s_{j+l+1})_{j,l=0}^{\infty} \succeq 0 \ . \tag{3.5}$$

Exercise 3.15. Prove Theorem 3.14.

Exercise 3.16. Let μ be a measure on $\mathbb{R}_+$ such that

$$\sum_{k\geq 0} s_k[\mu]^{-\frac{1}{2k}} = \infty \ . \tag{3.6}$$

Prove that μ is Stieltjes determinate, i.e. it shares its moments with no other measure on $\mathbb{R}_+$.

Remark 3.17. Beware: not every Stieltjes-determinate measure on $\mathbb{R}_+$ is Hamburger-determinate. Nevertheless, the conclusion of Exercise 3.16 can be strengthened (Wouk [1953]): a measure μ on $\mathbb{R}_+$ that satisfies (3.6) is necessarily Hamburger-determinate, i.e. it shares its moments with no other measure on $\mathbb{R}$.

The Hausdorff moment problem The Hausdorff moment problem concerns the moment sequences of measures that are supported on a bounded interval, e.g. $[0,1]$. Although there is a criterion similar to (3.5), the following, different, criterion is simpler and more convenient.

Theorem 3.18 (Hausdorff)**.** *A sequence (s_k) is the moment sequence of a measure μ supported on $\mathbb{R}_+$ if and only if*

$$\forall k, m \geq 0 \quad \sum_{j=0}^{m} (-1)^j \binom{m}{j} s_{j+k} \geq 0 \ . \tag{3.7}$$

In terms of the functional $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ sending $\lambda^k \mapsto s_j$, the condition (3.7) asserts that

$$\forall k, m \geq 0 \quad \Phi[\lambda^k (1-\lambda)^m] \geq 0 \ . \tag{3.8}$$

Proof of Hausdorff's theorem. The necessity of (3.8) is obvious. To prove sufficiency, first define Φ as above on $\mathbb{C}[\lambda]$. Introduce the Bernstein polynomial of a function $f : [0,1] \to \mathbb{C}$:

$$(B_N f)(\lambda) = \sum_{k=0}^{N} \binom{N}{k} f(k/N)\, \lambda^k (1-\lambda)^{N-k} \ .$$

Claim. If $R \in \mathbb{C}_{\leq n}[\lambda]$, then

$$B_N R = R + \sum_{j=1}^{n} \frac{E_j R}{N^j} \ , \quad E_j R \in \mathbb{C}[\lambda] \ . \tag{3.9}$$

Having the claim at hand, we let R be a polynomial which is non-negative on $[0,1]$, and estimate:

$$\Phi[R] = \Phi[B_N R] - \sum_{j=1}^{n} \frac{\Phi[E_j R]}{N^j} \ ;$$

the first term is non-negative, whereas the second term tends to zero as $N \to \infty$. Thus Φ is positive and can be extended to $C[0,1]$ using the M. Riesz theorem.

It remains to prove the claim. Let p, q be two formal variables. We shall later substitute $p = \lambda$, $q = 1 - \lambda$, but for now we keep them independent. Since

$$(p+q)^N = \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} \ ,$$

we have

$$(p\, \partial/\partial p)^n (p+q)^N = \sum_{k=0}^{N} \binom{N}{k} k^n p^k q^{N-k} = N^n \sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} (k/N)^n \ . \tag{3.10}$$

On the other hand, the "uncertainty relation"

$$(\partial/\partial p)p - p(\partial/\partial p) = \mathbb{1}$$

between the derivative and the operator of multiplication by p in $\mathbb{C}[p]$ implies that

$$(p\,\partial/\partial p)^n = \sum_{j=0}^{n} c_{j,n} p^j (\partial/\partial p)^j \ , \quad \text{where} \quad c_{n,n} = 1 \ .$$

Therefore

$$(p\,\partial/\partial p)^n (p+q)^N = \sum_{j=0}^{n} c_{j,n} N(N-1)\cdots(N-j+1) p^j (p+q)^{n-j} \ . \tag{3.11}$$

From the two representations (3.10) and (3.11)

$$\sum_{k=0}^{N} \binom{N}{k} p^k q^{N-k} (k/N)^n = N^{-n} \sum_{j=0}^{n} c_{j,n} N(N-1)\cdots(N-j+1) p^j (p+q)^{n-j} \ ,$$

whence, taking $p = \lambda$, $q = 1-\lambda$ as promised, we obtain the identity

$$\sum_{k=0}^{N} \binom{N}{k} \lambda^k (1-\lambda)^{N-k} (k/N)^n = \sum_{j=0}^{n} \frac{c_{j,n}}{N^{n-j}} (1-\frac{1}{N})(1-\frac{2}{N})\cdots(1-\frac{j-1}{N})\lambda^j \ ,$$

in which the left-hand side is $B_N[\lambda^n]$, whereas the right-hand side is brought to the form (3.9) by collecting powers of N. □

Remark 3.19. The polynomial $B_N f$ admits the following probabilistic interpretation: for $p \in [0,1]$,

$$(B_N f)(p) = \mathbb{E} f(\frac{1}{N}\sum_{j=1}^{N} \beta_j) \ ,$$

where β_j are independent Bernoulli random variables with $\mathbb{P}(\beta_j = 1) = p$.

Exercise 3.20 (S. Bernstein). If $f \in C[0,1]$, then $B_N f \rightrightarrows f$ as $N \to \infty$.

De Finetti theorem We briefly mention a theorem of De Finetti which follows from Hausdorff's theorem, and is in fact equivalent to it. See Diaconis [1987] and references therein for more details.

Definition 3.21. *A sequence $(X_1, X_2, X_3, \cdots)$ of random variables is called exchangeable if for any N and any permutation $\pi \in S_N$*

$$(X_1, \cdots, X_N) = (X_{\pi(1)}, \cdots, X_{\pi(N)}) \quad \text{in distribution.}$$

Example 3.22. *A sequence of independent identically distributed random variables is exchangeable, e.g.*

$$\mathbb{P}\left\{X_1 = \epsilon_1 \ , \cdots, X_N = \epsilon_N\right\} = \lambda^{\sum \epsilon_j} (1-\lambda)^{N - \sum \epsilon_j} \ , \quad \epsilon_1, \cdots, \epsilon_N \in \{0,1\}$$

(where $\lambda \in [0,1]$ is an arbitrary parameter).

Theorem 3.23 (De Finetti)**.** *Let* $(X_1, X_2, \cdots)$ *be an exchangeable sequence of random variables taking values in* $\{0,1\}$*. Then there exists a probability measure* μ *on* $[0,1]$ *so that*

$$\mathbb{P}\{X_1 = \epsilon_1 \ , \cdots , X_N = \epsilon_N\} = \int \lambda^{\sum \epsilon_j} (1-\lambda)^{N - \sum \epsilon_j} d\mu(\lambda) \ .$$

Proof. Consider the linear functional $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ which sends

$$\lambda^k (1-\lambda)^\ell \mapsto \mathbb{P}\{_1 = \epsilon_1 \ , \cdots , X_k = 1, X_{k+1} = 0 \ , \cdots , X_{k+\ell} = 0\} \ .$$

The definition is consistent (why?), and by Hausdorff's theorem there exists μ such that

$$\Phi[\lambda^k (1-\lambda)^\ell] = \int \lambda^k (1-\lambda)^\ell d\mu(\lambda) \ .$$

□

4 Orthogonal polynomials

Here we develop another approach to the Hamburger theorem, that will avoid functional analysis (the Riesz extension theorem) and provide more detailed information.

Let (s_k) be a moment sequence, i.e. a sequence such that

$$H = (s_{j+l})_{j,l=0}^\infty \succeq 0 \ . \tag{4.1}$$

Define a functional $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ by $\Phi[\sum a_j \lambda^j] = \sum a_j s_j$. By Hamburger's theorem Φ admits a realisation as integration with respect to a measure, however, we shall not use it (and in fact, we shall reprove Hamburger's theorem, without even using Proposition 3.2). Equivalently, we can start with a linear functional $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ which satisfies

$$\Phi[|Q|^2] \geq 0 \ , \quad Q \in \mathbb{C}[\lambda] \tag{4.2}$$

(as before, we denote $\bar{Q}(\lambda) = \sum \bar{\alpha}_j \lambda^j$ for $Q(\lambda) = \sum \alpha_j \lambda^j$, and $|Q|^2 = Q\bar{Q}$).

Define an inner product on $\mathbb{C}[\lambda]$ by $\langle P, Q \rangle = \Phi[P\bar{Q}]$. The condition (4.2) ensures that $\langle P, P \rangle \geq 0$. If H is not of finite rank, the monomials $1, \lambda, \lambda^2, \cdots$ are linearly independent. Thus $\langle P, P \rangle > 0$ whenever P is not identically zero. (The case of finite rank corresponds to measures supported on a finite number of points.)

Exercise 4.1. A positive-definite bilinear form $\langle \cdot, \cdot \rangle$ on $\mathbb{C}[\lambda]$ can be obtained via this construction if and only if

$$\langle \lambda P(\lambda), Q(\lambda) \rangle = \langle P(\lambda), \lambda Q(\lambda) \rangle \tag{4.3}$$

for any $P, Q \in \mathbb{C}[\lambda]$, i.e. multiplication by λ is a symmetric operator.

Applying the Gram–Schmidt orthonormalisation procedure, we obtain a sequence of polynomials P_k, where P_k is of degree k with positive leading coefficient, and

$$\langle P_k, P_\ell \rangle = \delta_{kl} \ .$$

If H is of finite rank, the same procedure yields a finite sequence of polynomials $P_0, \cdots, P_N$. We shall assume for the time being that H is not of finite rank, so that $\Phi[P] > 0$ whenever P is a non-negative polynomial which is not identically zero.

Definition 4.2. *The polynomials P_k are called the orthogonal polynomials with respect to the functional Φ.*

Example 4.3. $\operatorname{supp} \mu = [-1, 1]$, $d\mu/d\lambda = \frac{1}{\pi\sqrt{1-\lambda^2}}$. *Define*

$$T_k(\cos\theta) = \cos(k\theta) \ .$$

Note that this defines a polynomial of degree k with positive leading coefficient. Then:

$$\int T_k(\lambda) T_\ell(\lambda) d\mu(\lambda) = \int_0^\pi \cos(k\theta)\cos(\ell\theta) \frac{d\theta}{\pi} = \begin{cases} 1 \ , & k = \ell = 0 \\ \frac{1}{2} \ , & k = \ell > 0 \\ 0 \ , & k \neq \ell \ . \end{cases}$$

Thus $1, \sqrt{2}T_1, \sqrt{2}T_2, \sqrt{2}T_3, \cdots$ is the sequence of orthogonal polynomials with respect to μ. The polynomials T_k are called the Chebyshev polynomials of the first kind.

In the general case, the orthogonal polynomials admit the following formula.

Exercise 4.4. Prove that

$$P_k(\lambda) = \frac{1}{\sqrt{\det H_k \det H_{k+1}}} \det \begin{pmatrix} s_0 & s_1 & \cdots & s_k \\ s_1 & s_2 & \cdots & s_{k+1} \\ \cdots & & & \\ s_{k-1} & s_k & \cdots & s_{2k-1} \\ 1 & \lambda & \cdots & \lambda^k \end{pmatrix}$$

where $H_k = (s_{j+\ell})_{j,\ell=0}^{k-1}$.

Unfortunately, the computation of large determinants is as complicated (or as simple) as a Gram–Schmidt procedure.

4.1 Properties of orthogonal polynomials

We start with some elementary properties.

(a) $\langle P_k, Q \rangle = 0$ whenever $\deg Q < k$. (Obvious)

(b) P_k has k distinct real zeros.

Indeed, if $x_1 < \cdots < \lambda_m$ $(m < k)$ are the real zeros of odd multiplicity, then

$$P_k(\lambda)(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_m) \geq 0 \quad (\text{and not} \equiv 0)$$

and hence

$$\Phi[P_k(\lambda)(\lambda - \lambda_1)(\lambda - \lambda_2) \cdots (\lambda - \lambda_m)] > 0 \ .$$

However, the left-hand side should vanish by (a). □

(c) There exist coefficients a_k, b_k ($k \geq 0$) such that the following three-term recurrent relation holds:

$$\lambda P_k(\lambda) = b_k P_{k+1}(\lambda) + a_k P_k(\lambda) + b_{k-1} P_{k-1}(\lambda) \ , \quad k \geq 0 \ . \tag{4.4}$$

Indeed, $\lambda P_k(\lambda)$ is a linear combination of P_j ($0 \leq j \leq k+1$). For $j < k-1$,

$$\langle \lambda P_k(\lambda), P_j(\lambda) \rangle = \langle P_k(\lambda), \lambda P_j(\lambda) \rangle = 0$$

by (a). Finally,

$$\langle \lambda P_k(\lambda), P_{k+1}(\lambda) \rangle = \langle P_k(\lambda), \lambda P_{k+1}(\lambda) \rangle \ . \quad \square$$

Note that all b_k are positive (we still assume that H is not of finite rank).

Remark 4.5. As we see from the proof, the three-term recurrent relation is a consequence of the symmetry relation (4.3).

Let $\mathbb{C}_{\leq d}[\lambda] = \{P \in \mathbb{C}[\lambda] \mid \deg P \leq d\}$. Denote by $\pi_d : \mathbb{C}[\lambda] \to \mathbb{C}_{\leq d}[\lambda]$ the orthogonal projection.

(d) $\pi_d Q = \sum_{j=0}^{d} \langle Q, P_j \rangle P_j$ (obvious).

Denote:

$$K_d(z, z') = \sum_{j=0}^{d} P_j(z) P_j(\bar{z}') \ .$$

(e) (Reproducing property) For $Q \in \mathbb{C}_{\leq d}[\lambda]$, one has:

$$Q(z) = \Phi(K_d(z, \cdot) Q(\cdot)) = \langle K_d(z, \cdot), Q \rangle \ ,$$

and in particular

$$\Phi(K_d(z, \cdot) K_d(\cdot, z')) = \langle K_d(z, \cdot), K_d(z', \cdot) \rangle = K_d(z, z') \ .$$

Proposition 4.6 (Christoffel–Darboux formula)**.**

$$K_d(z, z') = b_d \, \frac{P_{d+1}(z) P_d(\bar{z}') - P_d(z) P_{d+1}(\bar{z}')}{z - \bar{z}'} \ .$$

Proof. It is sufficient to consider the case $z = \lambda, z' = \lambda' \in \mathbb{R}$. The expression

$$(\lambda - \lambda') K_d(\lambda, \lambda')$$

is a polynomial of degree $d+1$ in λ', hence it is a linear combination of $P_j(\lambda')$ ($0 \leq j \leq d+1$). Let us compute the coefficients: for $j \leq d-1$,

$$\langle (\lambda - \lambda') K_d(\lambda, \lambda'), P_j(\lambda') \rangle_{\lambda'} = \langle K_d(\lambda, \lambda'), (\lambda - \lambda') P_j(\lambda') \rangle_{\lambda'} = 0$$

by the reproducing property (e). Similarly (using (c))

$$\begin{aligned}
&\langle (\lambda - \lambda') K_d(\lambda, \lambda'), P_d(\lambda') \rangle_{\lambda'} \\
&= \langle K_d(\lambda, \lambda'), (\lambda - \lambda') P_d(\lambda') \rangle_{\lambda'} \\
&= \langle K_d(\lambda, \lambda'), \lambda P_d(\lambda') - (b_d P_{d+1}(\lambda') + a_d P_d(\lambda') + b_{d-1} P_{d-1}(\lambda')) \rangle_{\lambda'} \\
&= \lambda P_d(\lambda) - (a_d P_d(\lambda) + b_{d-1} P_{d-1}(\lambda)) = b_d P_{d+1}(\lambda)
\end{aligned}$$

and

$$\langle (\lambda - \lambda') K_d(\lambda, \lambda'), P_{d+1}(\lambda') \rangle_{\lambda'} = -b_d P_d(\lambda) \ .$$

Therefore

$$(\lambda - \lambda') K_d(\lambda, \lambda') = b_d (P_{d+1}(\lambda) P_d(\lambda') - P_d(\lambda) P_{d+1}(\lambda')) \ .$$

□

Exercise 4.7. Compute K_d for μ from Example 4.3.

4.2 Extremal problems

Denote by $\mathrm{LC}(Q)$ the leading coefficient of a polynomial Q. We still denote

$$\langle Q, R \rangle = \Phi(Q \bar{R}) \ , \quad \|Q\| = \sqrt{\Phi(|Q|^2)} \ .$$

Proposition 4.8.

$$\min_{\deg Q < d, \mathrm{LC}(Q) = 1} \|Q\| = \frac{1}{\mathrm{LC}(P_d)} \ ,$$

and the minimum is uniquely attained when $Q(\lambda) = P_d / \mathrm{LC}(P_d)$.

Proof. The minimum is attained when $Q(\lambda)$ is the projection of λ^d to the orthogonal complement of $\mathbb{C}_{\leq d-1}[\lambda]$, i.e.

$$Q(\lambda) = (\pi_d - \pi_{d-1})[\lambda^d] = \langle y^d, P_d(y) \rangle_y \, P_d(\lambda) = \frac{P_d(\lambda)}{\mathrm{LC}(P_d)} \ .$$

□

This can be generalised.

Proposition 4.9. *For any* $z \in \mathbb{C}$,

$$\min_{\deg Q \leq d, \ Q(z) = 1} \|Q(\lambda)\| = \frac{1}{\sqrt{K_d(z, z)}} \ ,$$

and the minimum is uniquely attained when $Q(\lambda) = \frac{K_d(\lambda, z)}{K_d(z, z)}$.

Proof. Let $Q = \sum_{j=0}^d c_j P_j$, then

$$1 = \sum c_j P_j(z) \le \sqrt{\sum |c_j|^2} \sqrt{\sum |P_j(z)|^2} = \sqrt{\sum |c_j|^2} \sqrt{K_d(z,z)} \ .$$

On the other hand, P_j form an orthonormal basis, hence $\sum |c_j|^2 = \|Q\|^2$. Thus

$$\|Q\|^2 \ge \frac{1}{\sqrt{K_d(z,z)}}$$

with equality attained when $c_j \propto \overline{P_j(z)}$, i.e.

$$Q(\lambda) = \frac{\sum_{j=0}^d P_j(\lambda)\overline{P_j(z)}}{\sum_{j=0}^d P_j(z)\overline{P_j(z)}} = \frac{K_d(\lambda,z)}{K_d(z,z)} \ .$$

□

4.3 Gaussian quadrature

Now we make a short digression. In a calculus course we have been taught the rectangle rule

$$\int_{-1}^1 f(\lambda) d\lambda \approx \frac{2}{2k+1} \sum_{j=-k}^{k} f(j/k)$$

and the trapezoid rule

$$\int_{-1}^1 f(\lambda) d\lambda \approx \frac{1}{2k} f(-1) + \frac{1}{k} \sum_{j=-k+1}^{k-1} f(j/k) + \frac{1}{2k} f(1) \ .$$

What is the optimal integration scheme? Following Gauss, we shall understand optimality as follows: we will look for a rule that is exact for all polynomials of a certain degree, and try to maximise this degree.

Example 4.10. *For $Q \in \mathbb{C}_{\le 5}[\lambda]$,*

$$\int_{-1}^1 Q(\lambda) d\lambda = \frac{5}{9} Q\left(-\sqrt{\frac{3}{5}}\right) + \frac{8}{9} Q(0) + \frac{5}{9} Q\left(\sqrt{\frac{3}{5}}\right) \ .$$

Note: there are 6 parameters and 6 degrees of freedom, so one can not do better.

Lemma 4.11 (Lagrange interpolation)**.** *Let $z_1, \cdots, z_d$ and $\alpha_1, \cdots, \alpha_d$ be complex numbers; assume that all the z_j are distinct. There is a unique polynomial $P \in \mathbb{C}_{\le d+1}[\lambda]$ such that*

$$P(z_j) = \alpha_j \quad (1 \le j \le d) \ ;$$

it is explicitly given by

$$P(z) = \sum_{j=1}^{d} \alpha_j \prod_{k \ne j} \frac{z - z_k}{z_j - z_k} = \sum_{j=}^{k} \alpha_j \frac{Q(z)}{Q'(z_j)(z - z_j)} \ ,$$

where $Q(z) = A \prod (z - z_j)$ (here $A \in \mathbb{C}$ is arbitrary).

Exercise 4.12. Prove the lemma.

Proposition 4.13 (Gaussian quadrature). *Let $\Phi : \mathbb{C}[\lambda] \to \mathbb{C}$ be a linear functional satisfying the positivity condition $\Phi[|R|^2] \geq 0$. Let P_j be the orthogonal polynomials constructed from Φ, and let $(\xi_j)_{j=1}^d$ be the zeros of P_d. Then*

$$\Phi[R] = \sum_{j=1}^{d} \Phi\left[\frac{P_d(\lambda)}{P_d'(\xi_j)(\lambda - \xi_j)}\right] R(\xi_j)$$

for any $R \in \mathbb{C}_{\leq 2d-1}[\lambda]$.

Proof. Let $R = P_d S_{d-1} + R_{d-1}$, where $\deg R_{d-1}, \deg S_{d-1} \leq d-1$. Then

$$\begin{aligned}
\Phi[R] &= \Phi[P_d S_{d-1}] + \Phi[R_{d-1}] = \Phi[R_{d-1}] \quad (\text{since } P_d \perp S_{d-1}) \\
&= \sum_{j=1}^{d} R_{d-1}(\xi_j) \Phi\left[\frac{P_d(\lambda)}{(\lambda - \xi_j)P_d'(\xi_j)}\right] \quad (\text{Lagrange interp.}) \\
&= \sum_{j=1}^{d} R(\xi_j) \Phi\left[\frac{P_d(\lambda)}{(\lambda - \xi_j)P_d'(\xi_j)}\right] \qquad (P_d(\xi_j) = 0)
\end{aligned}$$

□

The coefficients $\Phi\left[\frac{P_d(\lambda)}{P_d'(\xi_j)(\lambda-\xi_j)}\right]$ are called the Christoffel coefficients corresponding to Φ. It will be convenient to define the polynomials of the second kind:

$$Q_k(z) = \Phi_\lambda\left[\frac{P_k(z) - P_k(\lambda)}{z - \lambda}\right] \; ; \tag{4.5}$$

then the Christoffel coefficients can be rewritten as

$$\Phi\left[\frac{P_d(\lambda)}{P_d'(\xi_j)(\lambda - \xi_j)}\right] = \frac{Q_d(\xi_j)}{P_d'(\xi_j)} \; .$$

Exercise 4.14. Q_k is a polynomial of degree $k-1$; it satisfies the recurrence

$$zQ_k(z) = b_k Q_{k+1}(z) + a_k Q_k(z) + b_{k-1} Q_{k-1}(z) \; , \quad k \geq 1$$

with the initial conditions

$$Q_0(z) \equiv 0 \; , \; Q_1(z) = \frac{\sqrt{s_0}}{b_0} \; . \tag{4.6}$$

For comparison,

$$P_0(z) = \frac{1}{\sqrt{s_0}} \; , \quad P_1(z) = \frac{z - a_0}{b_0\sqrt{s_0}} \; .$$

Exercise 4.15. Prove the identity:

$$P_{k-1}Q_k - P_k Q_{k-1} \equiv \frac{1}{b_{k-1}} \quad (k \geq 1) \tag{4.7}$$

Proposition 4.16. *The Christoffel coefficients admit the alternate expression*

$$\Phi\left[\frac{P_d(\lambda)}{P_d'(\xi_j)(\lambda-\xi_j)}\right] = \frac{1}{K_{d-1}(\xi_j,\xi_j)} ,$$

and in particular are positive.

Proof. Let A_j be the j-th Christoffel coefficient, and let

$$L_j(\lambda) = \frac{P_d(\lambda)}{(\lambda-\xi_j)P_d'(\lambda)}$$

be the j-th interpolation polynomial. Then for any $c_1, \cdots, c_d \in \mathbb{C}$

$$\Phi[|\sum c_j L_j|^2] = \sum_{j,k} c_j \bar{c}_k \Phi[L_j \bar{L}_k] = \sum_j |c_j|^2 A_j .$$

In particular, $A_j \geq 0$ for all j. Since every $R \in \mathbb{C}_{\leq d-1}[\lambda]$ can be represented as

$$R = \sum R(\xi_j) L_j ,$$

we obtain

$$A_j = \min_{R\in\mathbb{C}_{\leq d-1}[\lambda], R(\xi_j)=1} \Phi[|R|^2] = \frac{1}{K_{d-1}(\xi_j,\xi_j)} .$$

□

Second proof of Hamburger's theorem. The Gaussian quadrature yields, for any $R \in \mathbb{C}_{\leq 2d-1}[\lambda]$:

$$\Phi[R] = \sum_{j=1}^{d} \frac{1}{K_{d-1}(\xi_j,\xi_j)} R(\xi_j) = \int R(\lambda) d\mu_d(\lambda) ,$$

where

$$\mu_d = \sum_{j=1}^{d} \frac{\delta_{\xi_j}}{K_{d-1}(\xi_j,\xi_j)} . \tag{4.8}$$

The sequence μ_d is precompact in weak topology (why?). For every $R \in \mathbb{C}[\lambda]$ and any limit point μ of the sequence (μ_d),

$$\Phi[R] = \int R\, d\mu$$

(note that R is not bounded, hence this also requires an additional argument.) Therefore μ is a solution of the moment problem. □

For the sequel, we need the following generalisation of the measure

$$\mu_d = \sum_{j=1}^{d} \frac{\delta_{\xi_j}}{K_{d-1}(\xi_j,\xi_j)} .$$

Let $\tau \in \mathbb{R}$ (sometimes it will be convenient to allow $\tau = \infty$ as well). Let

$$P_k^\tau = P_k - \tau P_{k-1} \ , \ Q_k^\tau = Q_k - \tau Q_{k-1} \ .$$

Then from (4.5)

$$Q_k^\tau(z) = \Phi_\lambda \left[\frac{P_k^\tau(z) - P_k^\tau(\lambda)}{z - \lambda} \right] \ . \tag{4.9}$$

Exercise 4.17. Prove that, for any $\tau \in \mathbb{R}$, the polynomial P_d^τ has d distinct real zeros.

Let ξ_j^τ be the zeros of P_d^τ. For $\tau = \infty$ we let ξ_j^τ be the zeros of P_{d-1}. Set

$$\mu_d^\tau = \sum_{j=1}^d \frac{Q_d^\tau(\xi_j^\tau)}{(P_d^\tau)'(\xi_j^\tau)} \delta_{\xi_j^\tau} \ , \quad \tau \in \bar{\mathbb{R}} \ . \tag{4.10}$$

Exercise 4.18. Prove that for any $\tau \in \bar{\mathbb{R}}$

$$\Phi[R] = \int R d\mu_d^\tau$$

for any $R \in \mathbb{C}_{\leq 2d-2}[\lambda]$. (Note that the maximal degree is in general one less than that of the optimal Gauss quadrature, corresponding to $\tau = 0$.)

Exercise 4.19. Prove that $\frac{Q_d^\tau(\xi_j^\tau)}{(P_d^\tau)'(\xi_j^\tau)} = \frac{1}{K_{d-1}(\xi_j^\tau, \xi_j^\tau)}$.

5 Description of solutions

5.1 Stieltjes transform

The Stieltjes transform of a finite measure μ is defined as

$$w(z) = w_\mu(z) = \int \frac{d\mu(\lambda)}{\lambda - z} \ , \quad z \in \mathbb{C} \setminus \mathbb{R} \ .$$

Exercise 5.1. For any $a < b$,

$$\lim_{\epsilon \to +0} \frac{1}{\pi} \int_a^b \Im w(\lambda + i\epsilon) \, d\lambda = \frac{\mu\{a\}}{2} + \mu(a, b) + \frac{\mu\{b\}}{2} \ ,$$

and in particular μ is uniquely determined by w.

If the moment sequence (s_k) is determinate, $w(z)$ is uniquely determined by (s_k). Otherwise, $w(z)$ may assume different values for different solutions. In this section, we shall study the following question: given a moment sequence (s_k) and a point $z \in \mathbb{C} \setminus \mathbb{R}$, describe the geometric locus of all the values of $w_\mu(z)$ as μ runs over the set of solutions of the moment problem.[13] Throughout this section, we assume that $s_0 = 1$.

[13]The original plan of Chebyshev was to describe the geometric locus of the numbers $\mu(-\infty, a]$ (for each fixed $a \in \mathbb{R}$). This is possible, but the use of the Stieltjes transform avoids complications such as that of Remark 5.4.

Define:

$$\mathcal{K}_d(z) = \{w_\mu(z) \mid \forall 0 \leq k \leq 2d-2 : \; s_k[\mu] = s_k\} \ , \quad \mathcal{K}(z) = \bigcap_{d \geq 1} \mathcal{K}_d(z) \ .$$

The goal in this section is to prove the following two results.

Theorem 5.2. *For each d and $z \in \mathbb{C} \setminus \mathbb{R}$, $\mathcal{K}_d(z)$ is a closed disc centred at*

$$\mathfrak{z}_d(z) = -\frac{Q_d(z)P_{d-1}(\bar{z}) - Q_{d-1}(z)P_d(\bar{z})}{P_d(z)P_{d-1}(\bar{z}) - P_{d-1}(z)P_d(\bar{z})}$$

and of radius

$$\mathfrak{r}_d(z) = \frac{1}{2|\Im z|} \frac{1}{K_{d-1}(z, z)} \ .$$

Hence $\mathcal{K}(z)$ is either a disc or a point. Let

$$\mathfrak{r}(z) = \lim_{d \to \infty} \mathfrak{r}_d(z) \ .$$

Theorem 5.3 (Invariability)**.** *If $\mathfrak{r}(z) > 0$ for some $z \in \mathbb{C} \setminus \mathbb{R}$, then $\mathfrak{r}(z) > 0$ for all $z \in \mathbb{C} \setminus \mathbb{R}$.*

That is, the moment problem is determinate if and only if

$$K(z, z) = \sum_{k=0}^{\infty} |P_k(z)|^2 = \infty$$

for some (or for any) $z \in \mathbb{C} \setminus \mathbb{R}$.

Remark 5.4. It is possible that a determinate measure μ has

$$K(\lambda, \lambda) < \infty$$

for some $\lambda \in \mathbb{R}$; in fact, this happens if and only if λ is an atom of μ.

5.2 The Stieltjes transform of the quadratures

We first describe the geometric locus of $w(z)$ as μ varies over the one-parametric family μ_d^τ of special solutions to the truncated moment problem

$$\forall R \in \mathbb{C}_{\leq 2q-2}[\lambda] \quad \Phi[R] = \int R(\lambda) d\mu(\lambda) \ . \tag{5.1}$$

Proposition 5.5. *The Stieltjes transform $w_d^\tau = w_{\mu_d^\tau}(z)$ is given by*

$$w_d^\tau(z) = -\frac{Q_d^\tau(z)}{P_d^\tau(z)} \ .$$

Proof. By Lagrange interpolation

$$\frac{Q_d^\tau(z)}{P_d^\tau(z)} = \sum_{j=1}^d \frac{Q_d^\tau(\xi_j^\tau)}{(z-\xi_j^\tau)(P_d^\tau)'(\xi_j^\tau)} = \int \frac{d\mu_d^\tau(\lambda)}{z-\lambda} \ .$$

□

Remark 5.6. Proposition 5.5 makes sense also for $\tau = \infty$.

Proposition 5.7. *Fix $z \in \mathbb{C} \setminus \mathbb{R}$. As τ varies in $\overline{\mathbb{R}}$, the quantity $w_d^\tau(z)$ describes the circle $\Gamma_d(z) = \{|w - \mathfrak{z}_d(z)| = \mathfrak{r}_d(z)\}$.*

Proof. We have:

$$w_d^\tau(z) = -\frac{Q_d^\tau(z)}{P_d^\tau(z)} = -\frac{Q_d^\tau(z) - \tau Q_{d-1}^\tau(z)}{P_d^\tau(z) - \tau P_{d-1}(z)}$$

This is a fractional-linear function of τ, hence the image of the (generalised) circle $\overline{\mathbb{R}}$ is a (generalised) circle. To find its equation, we rewrite

$$\begin{aligned} w_d^\tau(z) = &-\frac{Q_d(z)P_{d-1}(\bar{z}) - Q_{d-1}(z)P_d(\bar{z})}{P_d(z)P_{d-1}(\bar{z}) - P_{d-1}(z)P_d(\bar{z})} \\ &+ \frac{Q_d(z)P_{d-1}(z) - Q_{d-1}(z)P_d(z)}{P_d(z)P_{d-1}(z) - P_{d-1}(z)P_d(z)} \, \frac{P_d(\bar{z}) - \tau P_{d-1}(\bar{z})}{P_d(z) - \tau P_{d-1}(z)} \end{aligned}$$

and observe that

$$\left| \frac{P_d(\bar{z}) - \tau P_{d-1}(\bar{z})}{P_d(z) - \tau P_{d-1}(z)} \right| = 1 \ ,$$

whence the centre is at

$$\mathfrak{z}_d(z) = -\frac{Q_d(z)P_{d-1}(\bar{z}) - Q_{d-1}(z)P_d(\bar{z})}{P_d(z)P_{d-1}(\bar{z}) - P_{d-1}(z)P_d(\bar{z})}$$

whereas the radius is equal to

$$\mathfrak{r}_d(z) = \left| \frac{Q_d(z)P_{d-1}(z) - Q_{d-1}(z)P_d(z)}{P_d(z)P_{d-1}(z) - P_{d-1}(z)P_d(z)} \right| = \frac{1}{K_{d-1}(z)}$$

by (4.7) and the Christoffel–Darboux formula. □

5.3 Weyl circles

Recall that $\Gamma_d(z) = \{|w - \mathfrak{z}_d(z)| = \mathfrak{r}_d(z)\}$ and set $\mathcal{K}_d(z) = \operatorname{conv} \Gamma_d(z)$. We have shown that for the special solutions μ_d^τ of the truncated moment problem

$$s_k[\mu] = s_k \quad (0 \le k \le 2d-2) \tag{5.2}$$

the Stieltjes transform evaluated at the point $z \in \mathbb{C} \setminus \mathbb{R}$ lies on the circle $\Gamma_d(z)$. Now we prove:

Theorem 5.2. *For any solution μ of (5.2) and any $z \in \mathbb{C} \setminus \mathbb{R}$, $w_\mu(z) \in \mathcal{K}_d(z)$. Vice versa, for any $z \in \mathbb{C}\setminus\mathbb{R}$ and $w \in \mathcal{K}_d(z)$ there exists a solution μ of (5.2) with $w_\mu(z) = w$.*

Lemma 5.8. *For any $z \in \mathbb{C} \setminus \mathbb{R}$,*

$$\Gamma_d(z) = \left\{ w \mid \frac{\Im w}{\Im z} = \sum_{k=0}^{d-1} |wP_k(z) + Q_k(z)|^2 \right\} \tag{5.3}$$

$$\mathcal{K}_d(z) = \left\{ w \mid \frac{\Im w}{\Im z} \geq \sum_{k=0}^{d-1} |wP_k(z) + Q_k(z)|^2 \right\} . \tag{5.4}$$

Proof. Observe that $w \in \Gamma_d(z)$ if and only if

$$w = -\frac{Q_d^\tau(z) - \tau Q_{d-1}^\tau(z)}{P_d^\tau(z) - \tau P_{d-1}(z)} = \frac{\tau Q_{d-1}^\tau(z) - Q_d^\tau(z)}{-\tau P_{d-1}(z) + P_d^\tau(z)}$$

for some $\tau \in \overline{\mathbb{R}}$, i.e. if and only if

$$\frac{wP_d(z) + Q_d(z)}{wP_{d-1}(z) + Q_{d-1}(z)} \in \overline{\mathbb{R}} .$$

The last condition is equivalent to

$$(wP_d(z) + Q_d(z))(\bar{w}P_{d-1}(\bar{z}) + Q_{d-1}(\bar{z})) \in \mathbb{R} . \tag{5.5}$$

On the other hand,

$$\sum_{k=0}^{d-1} |wP_k(z) + Q_k(z)|^2 = \frac{\Im w}{\Im z} + 2ib_{d-1}\Im \left\{(wP_d(z) + Q_d(z))(\bar{w}P_{d-1}(\bar{z}) + Q_{d-1}(\bar{z}))\right\} . \tag{5.6}$$

This identity is a generalisation of the Christoffel–Darboux Formula (Proposition 4.6); it can be proved by the same method. We omit the proof, since we prove a more general fact in Corollary 7.2. Thus, w satisfies (5.5) if and only if

$$\sum_{k=0}^{d-1} |wP_k(z) + Q_k(z)|^2 = \frac{\Im w}{\Im z} .$$

We leave (5.4) as an exercise. □

Proof of Theorem 5.2. The "vice versa" part follows from Proposition 5.7 by taking convex combinations, therefore let us prove the first part. Consider an arbitrary solution μ of (5.2). The function $f(\lambda) = 1/(\lambda - z)$ lies in $L_2(\mu)$, and the polynomials $P_0, \cdots, P_{d-1}$ form an orthonormal system in this space. Therefore

$$\int |f(\lambda)|^2 d\mu(\lambda) \geq \sum_{k=0}^{d-1} \left| \int f(\lambda) P_k(\lambda) d\mu(\lambda) \right|^2 . \tag{5.7}$$

Observe that

$$|f(\lambda)|^2 = \frac{1}{\lambda - z}\frac{1}{\lambda - \bar{z}} = \frac{1}{\bar{z} - z}\left[\frac{1}{\lambda - \bar{z}} - \frac{1}{\lambda - z}\right] ,$$

hence

$$(\text{LHS of (5.7)}) = \frac{\Im w_\mu(z)}{\Im z} .$$

On the other hand,

$$\begin{aligned}\int f(\lambda)P_k(\lambda)d\mu(\lambda) &= \int \left[\frac{P_k(\lambda) - P_k(z)}{\lambda - z}d\mu(\lambda) + P_k(z)\int \frac{d\mu(\lambda)}{\lambda - z}\right] \\ &= Q_k(z) + w_\mu(z)P_k(z) ,\end{aligned}$$

hence

$$(\text{RHS of (5.7)}) = \sum_{k=0}^{d-1} |Q_k(z) + w_\mu(z)P_k(z)|^2 .$$

It remains to appeal to the second relation of Lemma 5.8. □

5.4 Invariability

We have a nested sequence of circles $\mathcal{K}_1(z) \supset \mathcal{K}_2(z) \supset \mathcal{K}_3(z) \supset \cdots$. From Lemma 5.8 and the equality $w_d^\infty(z) = w_{d-1}^0(z)$ we see that the boundaries $\Gamma_d(z)$ of $\mathcal{K}_d(z)$ and $\Gamma_{d-1}(z)$ of $K_{d-1}(z)$ intersect at a single point. Denote:

$$\mathcal{K}(z) = \bigcap_{d\geq 1} \mathcal{K}_d(z) .$$

This is either a disc or a point.

Proposition 5.9. *Let $z \in \mathbb{C} \setminus \mathbb{R}$, then $\mathcal{K}(z)$ is a disc if and only if*

$$\sum_{k=0}^{\infty} \left(|P_k(z)|^2 + |Q_k(z)|^2\right) < \infty . \tag{5.8}$$

Proof. For any $z \in \mathbb{C} \setminus \mathbb{R}$ there exists $w = \mathfrak{z}(z) \neq 0$ such that

$$\sum_{k=0}^{\infty} |wP_k(z) + Q_k(z)|^2 \leq \frac{\Im w}{\Im z} < \infty .$$

If $\mathcal{K}(z)$ is a disk, then also

$$\sum_{k=0}^{\infty} |P_k(z)|^2 < \infty \tag{5.9}$$

and hence

$$\sum_{k=0}^{\infty} |Q_k(z)|^2 < \infty .$$

Vice versa, (5.8) implies (5.9). □

Theorem 5.3 (Invariability)**.** *If (5.9) holds for some $z \in \mathbb{C} \setminus \mathbb{R}$, then it holds for all $z \in \mathbb{C}$. In particular, if $\mathcal{K}(z)$ is a disc for some $z \in \mathbb{C} \setminus \mathbb{R}$, then it is a disc for all $z \in \mathbb{C} \setminus \mathbb{R}$.*

Lemma 5.10. *Let $A = (a_{nk})_{n,k \geq 0}$ be an infinite matrix such that*

$$a_{nk} = 0 \quad (k \geq n) , \qquad \sum_{n,k} |a_{nk}|^2 < \infty .$$

Then there exists $C(A)$ such that, for any $y = (y_n)_{n \geq 0}$, the solution $x = (x_n)_{n \geq 0}$ of the system

$$y_n = x_n - \sum_{k=0}^{n-1} a_{nk} x_k$$

satisfies

$$\sum |x_n|^2 \leq C(A)^2 \sum |y_n|^2 . \tag{5.10}$$

In other words, $(\mathbb{1} - A)^{-1}$ is a bounded operator in ℓ_2.

Proof. It suffices to prove the relation (5.10) for vectors y of finite support, as long as the bound is uniform in the support. First, if $A_N = (a_{nk})_{n,k=0}^{N-1}$ is a finite matrix of dimension $N \times N$, then

$$x = (\mathbb{1} - A_N)^{-1} y = (\mathbb{1} + A_N + \cdots + A_N^{N-1}) y ,$$

hence

$$\|x\| \leq (1 + \|A_N\| + \cdots + \|A_N^{N-1}\|)\|y\| \leq C_1(A_N)\|y\| . \tag{5.11}$$

Now consider an infinite matrix A. Denote

$$N = \min \left\{ m \mid \sum_{n \geq m} \sum_{k=0}^{n-1} |a_{nk}|^2 < \frac{1}{8} \right\}$$

then by (5.11) applied to a principal submatrix A_N of A,

$$\sum_{n=0}^{N-1} |x_n|^2 \leq C_1(A_N)^2 \|y\| . \tag{5.12}$$

Fix y and denote

$$J(x) = \left\{ n \geq N \,\Big|\, |x_n| \leq 2 \sqrt{\sum_{k=0}^{n-1} |a_{nk}|^2} \; \|x\| \right\} .$$

Then

$$\sum_{n \in J(x)} |x_n|^2 \leq 4 \sum_{n \in J(x)} \sum_{k=0}^{n-1} |a_{nk}|^2 \|x\|^2 \leq 4 \sum_{n \geq N} \sum_{k=0}^{n-1} |a_{nk}|^2 \|x\|^2 \leq \frac{1}{2} \|x\|^2 . \tag{5.13}$$

On the other hand, for $n \geq N$, $n \notin J(x)$

$$|y_n| \geq |x_n| - \sum_{k=0}^{n-1} |a_{nk}||x_k| \geq |x_n| - \sqrt{\sum_{k=0}^{n-1} |a_{nk}|^2}\, \|x\| \geq \frac{1}{2}|x_n| \ ,$$

hence

$$\sum_{n \geq N,\, n \notin J(x)} |y_n|^2 \geq \frac{1}{4} \sum_{n \geq N,\, n \notin J(x)} |x_n|^2 \ . \tag{5.14}$$

From (5.13), (5.14) and (5.12)

$$\begin{aligned}\sum_n |x_n|^2 &= \sum_{n=0}^{N-1} + \sum_{n \geq N,\, n \notin J(x)} + \sum_{n \in J(x)} \leq \frac{3}{2} \left\{ \sum_{n=0}^{N-1} + \sum_{n \geq N,\, n \notin J(x)} \right\} \\ &\leq \frac{3}{2}(C(A_N)^2 + 4) \sum_n |y_n|^2 \ ,\end{aligned}$$

which proves that

$$\|(\mathbb{1} - A)^{-1}\| \leq \sqrt{\frac{3}{2}(C(A_N)^2 + 4)} \ .$$

□

Exercise 5.11. In the setting of Lemma 5.10, show that $\|(\mathbb{1} - A)^{-1}\|$ may be bounded by a quantity depending only on

$$\|A\|_{\mathrm{HS}}^2 = \sum_{k,n} |a_{k,n}|^2 \ .$$

Proof of Theorem 5.3. Suppose (5.9) holds for some $z_0 \in \mathbb{C} \setminus \mathbb{R}$. Then also

$$\sum |Q_n(z_0)|^2 < \infty$$

by Proposition 5.9. Observe that

$$\frac{P_d(z) - P_d(z_0)}{z - z_0} = \sum_{k=0}^{d-1} a_{d,k}(z_0) P_k(z) \ ,$$

where

$$\begin{aligned} a_{d,k}(z_0) &= \Phi\left[\frac{P_d(\lambda) - P_d(z_0)}{\lambda - z_0} P_k(\lambda)\right] \\ &= P_k(z_0)\Phi\left[\frac{P_d(\lambda) - P_d(z_0)}{\lambda - z_0}\right] + \Phi\left[\frac{P_d(\lambda) - P_d(z_0)}{\lambda - z_0}(P_k(\lambda) - P_k(z_0))\right] \\ &= P_k(z_0)Q_d(z_0) - P_d(z_0)Q_k(z_0) \ ; \end{aligned}$$

hence

$$\sum_{d=0}^{\infty}\sum_{k=0}^{d-1}|a_{d,k}(z_0)|^2 \leq 4\sum_k |P_k(z_0)|^2 \sum_d |Q_d(z_0)|^2 < \infty \ .$$

Invoking Lemma 5.10, we obtain that

$$\sum_{d=0}^{\infty} |P_d(z)|^2 < \infty \ .$$

□

Exercise 5.12. Under the assumptions of Theorem 5.3, the series

$$K(z,z') = \sum_{k=0}^{\infty} P_k(z) P_k(\bar{z}')$$

converges locally uniformly, and thus defines a function which is analytic in z and anti-analytic in z'.

6 Some applications

6.1 Completeness of polynomials*

Theorem 6.1 (Riesz). *Let μ be a solution of an indeterminate moment problem. The following are equivalent:*

1. *there exists $z \in \mathbb{C} \setminus \mathbb{R}$ such that $w_\mu(z) \in \Gamma(z) = \partial\mathcal{K}(z)$;*
2. *the relation $w_\mu(z) \in \Gamma(z)$ holds for any $z \in \mathbb{C} \setminus \mathbb{R}$;*
3. *polynomials are dense in $L_2(\mu)$.*

Proof. (3) $\Longrightarrow$ (2): if polynomials are dense in $L_2(\mu)$, then, in particular, $\phi_z(\lambda) = \frac{1}{\lambda - z}$ can be approximated by polynomials for any $z \in \mathbb{C} \setminus \mathbb{R}$, i.e.

$$\|\phi_z\|^2_{L_2(\mu)} = \sum_{k\geq 0} |\langle \phi_z, P_k\rangle|^2 \ .$$

Therefore $w_\mu \in \Gamma(z)$.

Since (2) $\Longrightarrow$ (1) is obvious, we proceed to (1) $\Longrightarrow$ (3). If $w_\mu(z) \in \Gamma(z)$, then ϕ_z and $\phi_{\bar{z}}$ can be approximated by polynomials. Next, the squares ϕ_z^2 and $\phi_{\bar{z}}^2$ lie in the closure of polynomials:

$$\begin{aligned}
&\int \left| \frac{1}{(\lambda - z)^2} - \frac{A}{\lambda - z} - \sum_{k=0}^{n} B_k \lambda^k \right|^2 d\mu(\lambda) \\
&\quad \leq \frac{1}{|\Im z|^2} \int \left| \frac{1}{\lambda - z} - A - \sum_{k=0}^{n} B_k \lambda^k (\lambda - z) \right|^2 d\mu(\lambda) \\
&\quad = \frac{1}{|\Im z|^2} \int \left| \frac{1}{\lambda - z} - (A + B_0) - \sum_{k=1}^{n} (B_{k-1} - zB_k)\lambda^k \right|^2 d\mu(\lambda) \ ,
\end{aligned}$$

and this expression can be made arbitrarily small by choosing suitable A and B_k. Proceeding in a similar fashion, we obtain that ϕ_z^k and $\phi_{\bar{z}}^k$ can be approximated by polynomials.

Suppose polynomials are not dense. Then there exists $0 \neq g \in L_2(\mu)$ such that $\langle g, f\rangle = 0$ for any f in the closure of polynomials. In particular,

$$\int \frac{g(\lambda)}{(\lambda - z)^k} d\mu(\lambda) = \int \frac{g(\lambda)}{(\lambda - \bar{z})^k} d\mu(\lambda) = 0 \ , \quad k = 1, 2, 3 \cdots$$

Let

$$u(\zeta) = \int \frac{g(\lambda)}{\lambda - \zeta} d\mu(\lambda) \ , \quad \zeta \in \mathbb{C} \setminus \mathbb{R} \ .$$

This function is analytic in $\mathbb{C}_\pm$, and

$$u(z) = u'(z) = u''(z) = \cdots = u(\bar{z}) = u'(\bar{z}) = u''(\bar{z}) = \cdots = 0 \ .$$

Hence $u \equiv 0$, and then $g \, d\mu = 0$ by the inversion formula for the Stieltjes transform, thus $g = 0$ in $L_2(\mu)$. □

Theorem 6.2 (Naimark)**.** *Consider the convex set of solutions to an indeterminate moment problem. A measure μ is an extreme point of this set if and only if polynomials are dense in $L_1(\mu)$.*

Proof. If μ is not an extreme point in the set of solutions, then it can be represented as $\mu = \alpha\mu_1 + (1 - \alpha)\mu_2$. Denote $\Phi[f] = \int f d\mu$ and $\Phi_1[f] = \int f d\mu_1$. Then we have:

$$|\Phi[f]| \leq \|f\|_{L_1(\mu)} \ , \quad |\Phi_1[f]| \leq \frac{1}{\alpha} \|f\|_{L_1(\mu)} \ ,$$

hence both functionals are continuous, and coincide on $\mathbb{C}[\lambda]$. Therefore polynomials are not dense.

Vice versa, if polynomials are not dense in $L_1(\mu)$, we can find a functional Φ_0 of norm 1 such that $\mathbb{C}[\lambda] \subset \operatorname{Ker} \Phi_0$. Let

$$\Phi_\pm[f] = \int f d\mu \pm \Phi_0[f] \ ,$$

then $\Phi_\pm$ are non-negative functionals. Therefore

$$\Phi_\pm[f] = \int f g_\pm d\mu$$

with $g_\pm \geq 0$. The measures $d\mu_\pm = g_\pm d\mu$ have the same moments as μ, and μ is their average. □

Definition 6.3. *A solution μ to a moment problem is called N-extreme if the equivalent conditions of Theorem 6.1 hold, and V-extreme if the equivalent conditions of Theorem 6.2 hold.*

Corollary 6.4. *Any N-extreme measure is V-extreme.*

Corollary 6.5. *For any $z \in \mathbb{C} \setminus \mathbb{R}$ and $w \in \Gamma(z)$, there exists a unique solution μ for which $w_\mu(z) = w$.*

Proof. If there were two such solutions μ_1 and μ_2, their average $\frac{\mu_1+\mu_2}{2}$ would also be N-extreme, in contradiction to Corollary 6.4. $\square$

Thus, in the indeterminate case the set of N-extreme solutions is homeomorphic to a circle.

6.2 Carleman's criterion, revisited*

Proposition 6.6 (Carleman)**.** *If $\sum b_k^{-1} = \infty$, then the moment problem is determinate.*

Proof. According to (4.7), $P_k Q_{k+1} - P_{k+1} Q_k = b_k^{-1}$, hence

$$\sum \frac{1}{b_k} = \sum (P_k(i) Q_{k+1}(i) - P_{k+1}(i) Q_k(i)) \leq \sum \left(|P_k(i)|^2 + |Q_k(i)|^2 \right) \ .$$

Thus the divergence of the left-hand side implies the divergence of the right-hand side and hence (by Proposition 5.9) the determinacy of the moment problem. $\square$

Now we can give another proof of

Corollary 2.12 (Carleman's criterion). *If $\sum_k s_{2k}^{-\frac{1}{2k}} = \infty$, the moment problem is determinate.*

Proof. Without loss of generality $s_0 = 1$. Then

$$\mathrm{LC}(P_k) = \frac{1}{b_0 b_1 \cdots b_{k-1}} \ ,$$

whence

$$b_0 \cdots b_{k-1} = b_0 \cdots b_{k-1} \Phi[|P_k|^2] = \Phi[\lambda^k P_k(\lambda)] \leq \sqrt{\Phi[\lambda^{2k}]} = \sqrt{s_{2k}} \ ,$$

hence by Carleman's inequality (2.6)

$$\sum_k s_{2k}^{-\frac{1}{2k}} \leq \sum_k (b_0 \cdots b_k)^{-\frac{1}{k}} \leq e \sum_k \frac{1}{b_k} \ ,$$

whence the divergence of the left-hand side implies determinacy by Proposition 6.6. $\square$

6.3 A condition for indeterminacy

Theorem 6.7 (Krein's condition)**.** *If μ has an absolutely continuous component with density $u(\lambda) = \mu'_{\mathrm{ac}}(\lambda)$ such that*

$$\int \log u(\lambda) \frac{d\lambda}{1+\lambda^2} > -\infty \ , \tag{6.1}$$

then μ is indeterminate.

Proof. The following proof is borrowed from Berg [2011]. A similar argument was used by Szegő in the context of the trigonometric moment problem.

It suffices to consider the case $d\mu = u(\lambda)d\lambda$. Recall that

$$K_d(z, z') = \sum_{k=0}^{d} P_k(z)P_k(\bar{z}') \ , \quad \int K_d(z, \lambda)K_d(\lambda, z')d\mu(\lambda) = K_d(z, z') \ ;$$

we shall prove that $\sup_d K_d(i, i) < \infty$.

Lemma 6.8. *If $K_d(z, z') = 0$, then either both z and z' are real, or $\Im z \Im z' < 0$.*

Proof. By the Christoffel–Darboux formula (Proposition 4.6), $K_d(z, z') = 0$ if and only if

$$\frac{P_{d+1}(z)}{P_d(z)} = \frac{P_{d+1}(\bar{z}')}{P_d(\bar{z}')} \ . \tag{6.2}$$

Now, the ratio P_{d+1}/P_d is only real on $\mathbb{R}$, since $P^\tau_{d+1} = P_{d+1} - \tau P_d$ has only real zeros for any $\tau \in \mathbb{R}$ (see Exercise 4.17). Also,

$$\lim_{y\to\infty} \frac{P_{d+1}(iy)}{P_d(iy)} = i\infty$$

since $\mathrm{LC}(P_{d+1}), \mathrm{LC}(P_d) > 0$. Therefore P_{d+1}/P_d maps $\mathbb{C}_+$ to $\mathbb{C}_+$ and $\mathbb{C}_-$ — to $\mathbb{C}_-$, and (6.2) is impossible if, say, $z \in \mathbb{C}_+$ and $z' \in \mathbb{C}_- \cup \mathbb{R}$. □

Exercise 6.9. If $R \in \mathbb{C}[z]$ is such that $R^{-1}(0) \subset \mathbb{C}_-$, then

$$\frac{1}{\pi} \int \frac{\log |R(\lambda)|}{1+\lambda^2} d\lambda = \log |R(i)| \ .$$

Now we conclude the proof of the theorem. Consider the integral

$$I = \int \log(|K_d(i, \lambda)|^2 u(\lambda)) \frac{d\lambda}{\pi(1+\lambda^2)} \ .$$

On the one hand, Jensen's inequality yields

$$I \leq \log \int |K_d(i, \lambda)|^2 u(\lambda) \frac{d\lambda}{\pi(1+\lambda^2)} \leq \log \int |K_d(i, \lambda)|^2 u(\lambda) \frac{d\lambda}{\pi} = \log \frac{|K_d(i, i)|}{\pi} \ .$$

On the other hand,

$$\begin{aligned} I &= 2\int \log|K_d(i,\lambda)|\frac{d\lambda}{\pi(1+\lambda^2)} + \int \log u(\lambda)\frac{d\lambda}{\pi(1+\lambda^2)} \\ &= 2\log|K_d(i,i)| + \int \log w(\lambda)\frac{d\lambda}{\pi(1+\lambda^2)} \ . \end{aligned}$$

Therefore

$$K_d(i,i) = |K_d(i,i)| \leq \frac{1}{\pi}\exp\left\{-\int \log u(\lambda)\frac{d\lambda}{\pi(1+\lambda^2)}\right\} \ .$$

□

Exercise 6.10 (A. Ostrowski). Let $d\mu(\lambda) = e^{-a(\lambda)}d\lambda$, where a is even, $a(0) > -\infty$, $t \mapsto a(e^t)$ is increasing and convex, and

$$\lim_{\lambda\to+\infty}\frac{a(\lambda)}{\log\lambda} = +\infty$$

(so that μ has finite moments). Then μ is determinate if and only if

$$\int \frac{a(\lambda)d\lambda}{1+\lambda^2} = \infty \ .$$

7 Jacobi matrices

Now we discuss the spectral theory of Jacobi matrices. We study the recurrence

$$(\star)_\lambda: \quad \lambda\psi(k) = b_k\psi(k+1) + a_k\psi(k) + b_{k-1}\psi(k-1) \ , \quad k \geq 1 \ , \tag{7.1}$$

where $a_k \in \mathbb{R}$ and $b_k \in \mathbb{R}$ are fixed coefficients, $\psi : \mathbb{Z}_+ \to \mathbb{C}$ is an unknown sequence, and λ is a parameter. We assume that $b_k > 0$ for all $k \geq 0$. The recurrence $(\star)_\lambda$ is a discrete analogue of a Sturm–Liouville equation

$$-\frac{d}{dx}\left[p(x)\frac{d\psi(x)}{dx}\right] + q(x)\psi(x) = \lambda w(x)\psi(x) \ .$$

We state an important property of solutions:

Proposition 7.1. *Let ψ_j $(j = 1,2)$ be a solution of $(\star)_{\lambda_j}$. Then for any $k \geq 1$*

$$b_k\begin{vmatrix} \psi_1(k) & \psi_2(k) \\ \psi_1(k+1) & \psi_2(k+1) \end{vmatrix} - b_{k-1}\begin{vmatrix} \psi_1(k-1) & \psi_2(k-1) \\ \psi_1(k) & \psi_2(k) \end{vmatrix} = (\lambda_2-\lambda_1)\psi_1(k)\psi_2(k) \ .$$

The formula remains true for complex λ_j.

Proof. The left-hand side is equal to

$$\begin{vmatrix} \psi_1(k) & \psi_2(k) \\ b_k\psi_1(k+1) + b_{k-1}\psi_1(k-1) & b_k\psi_2(k+1) + b_{k-1}\psi_2(k-1) \end{vmatrix}$$
$$= \begin{vmatrix} \psi_1(k) & \psi_2(k) \\ (\lambda_1 - a_k)\psi_1(k) & (\lambda_2 - a_k)\psi_2(k) \end{vmatrix} = \begin{vmatrix} \psi_1(k) & \psi_2(k) \\ \lambda_1\psi_1(k) & \lambda_2\psi_2(k) \end{vmatrix} .$$

□

Corollary 7.2 (Discrete Green formula)**.** *Under the assumptions of Proposition 7.1,*

$$\sum_{k=1}^{d} \psi_1(k)\psi_2(k) = \frac{1}{\lambda_2 - \lambda_1} \left\{ b_d \begin{vmatrix} \psi_1(d) & \psi_2(d) \\ \psi_1(d+1) & \psi_2(d+1) \end{vmatrix} - b_0 \begin{vmatrix} \psi_1(0) & \psi_2(0) \\ \psi_1(1) & \psi_2(1) \end{vmatrix} \right\} .$$

Remark 7.3. The Christoffel–Darboux formula (Proposition 4.6) as well as the formula (5.6) are special cases of Corollary 7.2, whereas (4.7) follows directly from Proposition 7.1.

7.1 Connection to spectral theory

Define a semiinfinite matrix

$$J = \begin{pmatrix} a_0 & b_0 & 0 & 0 & 0 & \cdots \\ b_0 & a_1 & b_1 & 0 & 0 & \cdots \\ 0 & b_1 & a_1 & b_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \end{pmatrix}$$

It is called a Jacobi matrix. Then $(\star)_\lambda$ with the initial condition

$$\lambda\psi(0) = b_0\psi(1) + a_0\psi(0)$$

looks like an eigenvector equation: $J\psi = \lambda\psi$, and λ becomes the spectral parameter. The matrix J is symmetric, so we would expect a nice spectral theorem such as in Section 3.2. We would of course like the spectrum to be real. However:

Exercise 7.4. Let $a_k = 0$ and $b_k = k$! Prove that, for any $z \in \mathbb{C}\backslash\mathbb{R}$, there exists a solution $\psi \in \ell_2$ to $J\psi = z\psi$.

This is a sign that we have not imposed proper boundary conditions at infinity. Indeed, the problem for a finite interval:

$$\lambda\psi(k) = b_k\psi(k+1) + a_k\psi(k) + b_{k-1}\psi(k-1) \ , \quad 1 \le k \le d-1$$

is well-posed only with boundary conditions at both edges. The eigenvalues of

$$J_{d+1} = \begin{pmatrix} a_0 & b_0 & 0 & 0 & 0 & \cdots \\ b_0 & a_1 & b_1 & 0 & 0 & \cdots \\ 0 & b_1 & a_1 & b_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ & & & b_{d-2} & a_{d-1} & b_{d-1} \\ 0 & 0 & \ddots & 0 & b_{d-1} & a_d \end{pmatrix}$$

correspond to the Dirichlet boundary conditions

$$\lambda\psi(0) = b_0\psi(1) + a_0\psi(0) \ , \tag{7.2}$$

$$\lambda\psi(d) = a_d\psi(d) + b_{d-1}\psi(d-1) \ . \tag{7.3}$$

However, a_d may be chosen arbitrarily, therefore there is in fact a one-parametric family of good (self-adjoint) boundary conditions corresponding to different choices of a_d. If the sequences (a_k) and (b_k) are bounded, the matrix J defines a bounded self-adjoint operator with real spectrum, hence a situation such as in Exercise 7.4 can not occur. That is, in this case there is no need for any boundary conditions at infinity. As we shall see, this is a reflection of the following fact: the solutions corresponding to different choices of a_d coalesce in the limit $d \to \infty$.

We turn to the following question: given (a_k) and (b_k), when is there need for boundary conditions at infinity? To answer this question, consider the matrices

$$J_{d+1}(\tau) = \begin{pmatrix} a_0 & b_0 & 0 & 0 & 0 & \cdots \\ b_0 & a_1 & b_1 & 0 & 0 & \cdots \\ 0 & b_1 & a_1 & b_2 & 0 & \cdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & \\ & & & b_{d-2} & a_{d-1} & b_{d-1} \\ 0 & 0 & \ddots & 0 & b_{d-1} & a_d - \tau b_d \end{pmatrix} \ , \quad \tau \in \mathbb{R} \ .$$

The spectral measure of $J_{d+1}(\tau)$ corresponding to the vector δ_0 is given by

$$\mu^\tau_{d+1} = \sum_{j=1}^{d+1} |\psi^\tau_j(0)|^2 \delta_{\xi^\tau_j} \ , \quad \mu^\tau_{d+1}(B) = \sum_{\xi^\tau_j \in B} |\psi^\tau_j(0)|^2 \ , \tag{7.4}$$

where ξ^τ_j are the eigenvalues of $J_d(\tau)$, and ψ^τ_j are the associated eigenvalues. The similarity in notation to (4.10) is of course not coincidental.

Definition 7.5. *We say that a Jacobi matrix J is of type* **D** *if there exists a probability measure μ such that for any choice of $\tau_d \in \mathbb{R}$*

$$\mu^{\tau_d}_d \underset{d\to\infty}{\longrightarrow} \mu \quad \textit{in weak topology} \ .$$

Otherwise, we say that J is of type **C**.

7.2 Connection to the moment problem

In the sequel we assume that $b_k > 0$. This does not entail a loss of generality (why?).

Theorem 7.6 (Perron)**.** *For every recurrence of the form $(\star)_\lambda$ with $b_k > 0$ there exists a unique moment sequence $s_0 = 1, s_1, \cdots$ such that the corresponding orthogonal polynomials $P_k(\lambda)$ satisfy $(\star)_\lambda$ with the initial conditions*

$$P_0(\lambda) \equiv 1, \quad P_1(\lambda) = \frac{\lambda - a_0}{b_0} \ . \tag{7.5}$$

Proof. Define a sequence of polynomials $P_k(\lambda)$ that satisfy $(\star)_\lambda$, with the initial conditions (7.5). This sequence forms a (linear-algebraic) basis in $\mathbb{C}[\lambda]$, and $\mathrm{LC}(P_k) > 0$. Define an inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{C}[\lambda]$:

$$\langle \sum_j c_j P_j, \sum_k c'_k P_k \rangle = \sum_j c_j \bar{c}'_j \ .$$

It is obviously positive-definite. Let us show that there is a (unique) sequence (s_k) such that

$$\langle \sum c_j \lambda^j, \sum c'_k \lambda^k \rangle = \sum_{j,k} c_j \bar{c}'_k s_{j+k} \ .$$

By Exercise 4.1, it suffices to show that

$$\langle \lambda R(\lambda), S(\lambda) \rangle = \langle R(\lambda), \lambda S(\lambda) \rangle$$

for any pair of polynomials R, S. It suffices to consider the case $R = P_j$, $S = P_k$. Then

$$\begin{aligned} \langle \lambda P_j(\lambda), P_k(\lambda) \rangle &= \langle b_j P_{j+1} + a_j P_j + b_{j-1} P_{j-1}, P_k \rangle \\ &= b_j \delta_{j+1,k} + a_j \delta_{j,k} + b_{j-1} \delta_{j-1,k} = b_{k-1} \delta_{j,k-1} + a_k \delta_{j,k} + b_k \delta_{j,k+1} \\ &= \langle P_j, b_{k-1} P_{k-1} + a_k P_k + b_k P_{k+1} = \langle P_j(\lambda), \lambda P_k(\lambda) \rangle \ . \end{aligned}$$

The sequence (s_k) which we have constructed is a moment sequence, and the associated orthogonal polynomials are exactly the polynomials P_k. □

Theorem 7.7. *Let J be a Jacobi matrix with $b_k > 0$. Then J is of type* **D** *if and only if the corresponding moment problem is determinate.*

Lemma 7.8. $\det(z - J_d(\tau)) = b_0 \cdots b_{d-1} P_d^\tau(z)$.

Proof. It suffices to prove the proposition for $\tau = 0$, since

$$P_d^\tau = P_d - \tau P_{d-1} \ , \quad \det(J_d(\tau) - z) = \det(J_d - z) - \tau b_{d-1} \det(J_{d-1} - z) \ .$$

For $d = 1$ the identity follows from (7.5). The induction step follows from the recurrent relation $(\star)_z$. □

Lemma 7.9. *Let J be a Jacobi matrix with $b_k > 0$. The measures μ_d^τ defined in (7.4) coincide with those defined in (4.10) for the associated moment problem.*

Proof. In this proof, let us denote the measure defined in (7.4) by $\widetilde{\mu}_d^\tau$. Observe that for $0 \leq k \leq 2d - 2$

$$s_k[\widetilde{\mu}_d^\tau] = (J_d(\tau)^k)_{00} = (J^k)_{00} = \Phi[\lambda^k] \ ;$$

using this equality for $0 \leq k \leq d - 1$ completes the proof. □

Proof of Theorem 7.7. A sequence of probability measures (ν_d) is weakly convergent if and only if the sequence of Stieltjes transforms converges pointwise in $\mathbb{C} \setminus \mathbb{R}$. If the moment problem is determinate, the sequence (μ_d^τ) converges to a τ-independent limit, therefore the geometric locus of the values of $w_d^\tau(z)$ shrinks to a point for every $z \in \mathbb{C}\setminus\mathbb{R}$. Hence J is of type **D**.

Vice versa, if J is of type **D**, the geometric locus of the values of $w_\mu(z)$ as μ ranges over the collection of solutions to the moment problem degenerates to a point for any $z \in \mathbb{C} \setminus \mathbb{R}$, hence the moment problem is determinate. □

7.3 Inclusion in the general framework

Let J be a Jacobi matrix with $b_k > 0$. It defines an operator acting on the space of finitely-supported sequences in ℓ_2; this operator is symmetric:

$$\langle Ju, v\rangle = \langle u, Jv\rangle \tag{7.6}$$

for such u, v.

Recall several constructions and facts from spectral theory (see Akhiezer and Glazman [1993]). Let T be an operator defined on a dense subspace D of a Hilbert space $\mathcal{H}$. Consider the set of pairs

$$\Gamma = \left\{(u, u') \in \mathcal{H} \times \mathcal{H} \,\middle|\, \exists u_n \in D : u_n \to u \ , \ Tu_n \to u'\right\} \ .$$

If T is symmetric, i.e.

$$\langle Tu, v\rangle = \langle u, Tv\rangle \ ,$$

then Γ is a graph of an operator $\bar{T}$ which is called the closure of T. Also consider the collection

$$\Gamma^* = \left\{(v, v') \in \mathcal{H} \times \mathcal{H} \,\middle|\, \forall u \in D \ \langle Tu, v\rangle = \langle u, v'\rangle\right\} \ .$$

Without any assumptions, Γ^* is the graph of an operator T^*, called the adjoint operator; it is always closed (i.e. its closure coincides with itself). If T is symmetric, $\bar{T}^* = T^*$ and $T^{**} = \bar{T}$. A symmetric operator T is called self-adjoint if $T^* = T$ and essentially self-adjoint if $T^* = \bar{T}$.

Self-adjoint operators admit a spectral theorem similar to Theorem 3.9. On the other hand, symmetric operators that are not self-adjoint do not admit a natural spectral decomposition.

Let T be a symmetric operator defined on a dense subspace D of a Hilbert space $\mathcal{H}$. The invariability theorem asserts that the functions

$$z \mapsto \dim \mathrm{Ker}(T^* - z)$$

are constant in $\mathbb{C}_+$ and in $\mathbb{C}_-$. The numbers $n_\pm(T) = \dim \mathrm{Ker}(T^* \mp i)$ are called the deficiency indices of T. The operator T is self-adjoint if and only if its deficiency indices are $(0, 0)$. The operator T admits a self-adjoint extension $\widetilde{T} : \mathcal{H} \to \mathcal{H}$ if and only if its deficiency indices are equal. If the deficiency indices are $(n_+(T), n_-(T))$, the self-adjoint extensions are parametrised by unitary bijections

$$U : \mathrm{Ker}(T^* - i) \longleftrightarrow \mathrm{Ker}(T^* + i) \ .$$

Theorem 7.10. *Let J be a Jacobi matrix with $b_k > 0$. If J is of type* $\mathbf{D}$*, the corresponding operator J is essentially self-adjoint. If J is of type* $\mathbf{C}$*, the deficiency indices of the corresponding operator are* $(1,1)$*. The spectral measures of the self-adjoint extensions (with respect to the vector δ_0) are exactly the N-extreme solutions (Definition 6.3) to the moment problem.*

Exercise 7.11. Let $\mathcal{H}$ be a separable Hilbert space, and let $T : \mathcal{H} \to \mathcal{H}$ be a bounded self-adjoint operator with a cyclic vector v_1. There exists an orthonormal basis $(v_j)_{j=1}^{\infty}$ of $\mathcal{H}$ in which T is represented by a Jacobi matrix.

8 The multidimensional moment problem*

The two-dimensional counterpart of the map $\mathcal{S}$ from the introduction takes a measure μ on $\mathbb{R}^2$ to the array $(s_{\ell,k}[\mu])_{\ell,k\geq 0}$ of its moments

$$s_{\ell,k}[\mu] = \int \lambda_1^\ell \lambda_2^k d\mu(\lambda) \ .$$

Similarly to the one-dimensional case, any element of the image satisfies

$$\sum_{\ell,\ell',k,k'\geq 0} s_{\ell+\ell',k+k'} z_{\ell,k} \bar{z}_{\ell',k'} \geq 0 \ , \quad (z_{\ell,k}) \in \mathbb{C}^{\mathbb{Z}_+^2} \ . \tag{8.1}$$

However, unlike the one-dimensional case, this condition is necessary but not sufficient.

8.1 Existence

Proposition 8.1 (Hilbert)**.** *There exists $P \in \mathbb{R}[\lambda, \kappa]$ which can not be represented as a sum of squares.*

Hilbert's proof was not constructive; the first example (which we reproduce below) was constructed in the 1960-s by Motzkin [1967].

Proof. Let

$$M(\lambda, \kappa) = \lambda^4 \kappa^2 + \lambda^2 \kappa^4 + 1 - 3\lambda^2 \kappa^2 \ .$$

By the arithmetic-geometric mean inequality

$$M(\lambda, \kappa) \geq 3\sqrt[3]{\lambda^6 \kappa^6} - 3\lambda^2\kappa^2 = 0 \ .$$

Assume that M is a sum of squares: $M = \sum_j h_j^2$, $h_j \in \mathbb{R}[\lambda_1, \lambda_2]$. Denote the coefficients of h_j as follows:

$$\begin{aligned} h_j(\lambda, \kappa) = A_j\lambda^3 + B_j\lambda^2\kappa + C_j\lambda\kappa^2 + D_j\kappa^3 \\ + E_j\lambda^2 + F_j\lambda\kappa + G_j\kappa^2 \\ + H_j\lambda + I_j\kappa \\ + J_j \ . \end{aligned}$$

Comparing the highest-degree coefficients, we consecutively prove that

$$A_k = D_k = 0 \ , \quad E_k = G_k = 0 \ , \quad H_k = I_k = 0 \ .$$

From the equalities $M(\pm 1, \pm 1) = 0$ we obtain that $h_k(\pm 1, \pm 1) = 0$, i.e.

$$B_k + C_k + F_k + J_k = B_k - C_k - F_k + J_k = -B_k + C_k - F_k + J_k = -B_k - C_k + F_k + J_k = 0$$

whence $B_k = C_k = F_k = J_k = 0$. □

Using a Hahn–Banach argument in an appropriate topology, this can be shown to imply (see Berg [1987] and references therein):

Theorem 8.2. *There exists an array $(s_{\ell,k})$ satisfying (8.1) which is not a moment array, i.e. it does not admit a representation*

$$s_{\ell,k} = s_{\ell,k}[\mu] \ , \quad \ell, k \geq 0$$

with a positive measure μ.

See also Friedrich [1985] for an explicit construction of such an array $(s_{\ell,k})$.

Thus, Hamburger's theorem fails in two dimensions. On the other hand, Hausdorff's theorem can be extended:

Exercise 8.3. Let $\Phi : \mathbb{C}[\lambda, \kappa] \to \mathbb{C}$. The following are equivalent:

1. there exists a measure μ supported on the simplex $\{\lambda, \kappa \geq 0, \lambda + \kappa \leq 1\}$ such that

$$\Phi[R] = \int R \, d\mu \ , \quad R \in \mathbb{C}[\lambda, \kappa] \ ;$$

2. $\Phi[\lambda^\ell \kappa^k (1 - \lambda - \kappa)^m] \geq 0$ for any $\ell, m, k \geq 0$.

More generally, let $K \subset \mathbb{R}^2$ be a convex body (a compact convex set with non-empty interior); then

$$K = \bigcap_{\chi \in K^\circ} \left\{ \lambda \in \mathbb{R}^2 \, \middle| \, \langle \xi, \lambda \rangle \leq 1 \right\} \ ,$$

where

$$K^\circ = \left\{ \xi \in \mathbb{R}^2 \, \middle| \, \forall \lambda \in K : \langle \xi, \lambda \rangle \leq 1 \right\} \ .$$

Exercise 8.4 (Maserick). Let $\Phi : \mathbb{C}[\lambda, \kappa] \to \mathbb{C}$. The following are equivalent:

1. there exists a measure μ supported on K such that

$$\Phi[R] = \int R \, d\mu \ , \quad R \in \mathbb{C}[\lambda, \kappa] \ ;$$

2. $\Phi\left[R^2 \prod_{\xi \in \Xi} \langle \xi, \lambda \rangle\right] \geq 0$ for any finite $\Xi \subset K^\circ$ and any $R \in \mathbb{C}[\lambda, \kappa]$.

Remark 8.5. The following generalisation, due to Schmüdgen [1991], requires arguments from real algebraic geometry. Let $\mathfrak{P} \subset \mathbb{R}[\lambda, \kappa]$, and let

$$K = \bigcap_{P \in \mathfrak{P}} \{P \geq 0\} \ .$$

Let $\Phi : \mathbb{C}[\lambda, \kappa] \to \mathbb{C}$. If K is compact, the following are equivalent:

1. there exists a measure μ supported on K such that

$$\Phi[R] = \int R \, d\mu \ , \quad R \in \mathbb{C}[\lambda, \kappa] \ ;$$

2. $\Phi\left[R^2 \prod_{P \in \Xi} P\right] \geq 0$ for any finite $\Xi \subset \mathfrak{P}$ and any $R \in \mathbb{C}[\lambda, \kappa]$.

8.2 Carleman's criterion

Exercise 8.6. Let $(s_{\ell,k})$ be a moment array. If

$$\sum_{\ell=0}^{\infty} s_{2\ell,0}^{-\frac{1}{2\ell}} = \sum_{k=0}^{\infty} s_{0,2k}^{-\frac{1}{2k}} = \infty \ ,$$

then the corresponding moment problem is determinate.

The following theorem, due to Nussbaum [1965], is more surprising.

Theorem 8.7 (Nussbaum). *Let $(s_{\ell,k})$ be an array satisfying (8.1), and such that*

$$\sum_{\ell=0}^{\infty} s_{2\ell,0}^{-\frac{1}{2\ell}} = \infty \ .$$

Then $(s_{\ell,k})$ is a moment array.

Proof. Let $\Phi : \mathbb{C}[\lambda, \kappa] \to \mathbb{C}$ be the linear functional sending $\lambda^\ell \kappa^k$ to $s_{\ell,k}$. For any $p \in \mathbb{C}[\lambda]$ and $q \in \mathbb{C}[\kappa]$,

$$\Phi\left[p(\lambda)^2 q(\kappa)^2\right] \geq 0 \ ,$$

hence for any p there exists a measure $\tau[p^2]$ such that

$$\Phi\left[p(\lambda)^2 q(\kappa)\right] = \int q(\kappa) d\tau[p^2](\kappa) \ , \quad q \in \mathbb{C}[\kappa] \ .$$

Observe that

$$\Phi[p(\lambda)^2 \kappa^{2k}] \leq \sqrt{\Phi[p(\lambda)^4] \Phi[\kappa^{4k}]} \ ,$$

hence

$$\sum_k \Phi[p(\lambda)^2 \kappa^{2k}]^{-\frac{1}{2k}} \geq C_p \sum_k \Phi[\kappa^{4k}]^{-\frac{1}{4k}} = \infty \ .$$

Therefore $\tau[p^2]$ is defined uniquely. This crucial property allows us to define $\tau[p]$ for all polynomials $p \in \mathbb{C}[\lambda]$. Indeed, any such p can be represented as

$$p = p_1^2 + p_2^2 - p_3^2 - p_4^2 ,$$

and, if we set

$$\tau[p] = \tau[p_1^2] + \tau[p_2^2] - \tau[p_3^2] - \tau[p_4^2] ,$$

the value of this expression does not depend on the choice of the decomposition, and, moreover, depends linearly on p (why?) Consequently,

$$\Phi[p(\lambda)q(\kappa)] = \int q(\kappa)\, d\tau[p](\kappa) .$$

Next, $\tau[p]$ is monotone in p: if $p_1 \leq p_2$ on $\mathbb{R}$, then $\tau[p_1] \leq \tau[p_2]$. Indeed, $p_2 - p_1$ is a sum of squares, and by construction τ associates a positive measure to each square. Hence for every $B \subset \mathbb{R}$ there exists a measure σ_B such that

$$\tau[p](B) = \int p(\lambda) d\sigma_B(\lambda) . \tag{8.2}$$

We apply this as follows: on the j-th step, construct $\sigma_B^{(j)}$ which satisfies (8.2) for B of the form $[i/2^j, (i+1)/2^j)$. If B is a union of such elementary intervals, we define $\sigma_B^{(j)}$ as the sum of the corresponding measures. If B is a union of dyadic intervals, then, for sufficiently large j, $\sigma_B^{(j)}$ is a solution to the same moment problem (8.2). The set of such solutions is precompact in weak topology; choose a sequence $(j_r)_{r \geq 1}$ such that the sequence $(\sigma_B^{(j_r)})$ converges for any (dyadic) B:

$$\sigma_B^{(j_r)} \to \sigma_B .$$

Then σ_B is again a solution to the moment problem (8.2), and is, by construction, monotone non-decreasing as a function of B. Denote

$$M(\lambda, \kappa) = \lim_{\kappa' \to \kappa + 0} \sigma_{(-\infty, \kappa']}(-\infty, \lambda] ,$$

where the limit is taken along $\kappa' \in \bigcup_{j \geq 0} 2^{-j}\mathbb{Z}$. Then M is monotone non-decreasing in both λ and κ, and

$$\Phi[p(\lambda)q(\kappa)] = \int p(\lambda)q(\kappa) dM(\lambda, \kappa) .$$

Thus dM is the requested solution to the moment problem

$$s_{\ell,k}[\mu] = s_{\ell,k} , \quad \ell, k \geq 0 . \qquad \square$$

Review questions (Please send me the solutions to a few of them before 11.1.2018)

1. For which $\alpha \in \mathbb{R}$ is the measure $e^{-|\lambda| \log^\alpha(|\lambda|+e)} d\lambda$ determinate?

2. Let $\mathcal{M}$ be a log-convex sequence of positive numbers satisfying (2.4). The class of C^∞ functions f such that (2.2) holds for all even k is quasianalytic.

3. Prove that the polynomials $(P_k)_{k=0}^N$ defined via $P_0 \equiv 1$, $P_1(\lambda) = \frac{1}{\sqrt{N}}\lambda$ and

$$\lambda P_k(\lambda) = \sqrt{(N-k)(k+1)}\, P_{k+1}(\lambda) + \sqrt{(N-k+1)k}\, P_{k-1}(\lambda)\, (1 \le k \le N-1)$$

   are orthogonal with respect to the measure $\mu_N = \frac{1}{2^N} \sum_{j=0}^N \binom{N}{j} \delta_{N-2j}$.

4. Let μ be a determinate measure. Prove that $\lim_{k\to\infty} \frac{Q_k(z)}{P_k(z)} = -w_\mu(z)$ for each $z \in \mathbb{C}\backslash\mathbb{R}$.

5. Suppose $f \in L_2(\mu)$ lies in the closure of polynomials in $L_2(\mu)$. If μ is indeterminate, there exists an entire function which coincides μ-almost-everywhere with f.

6. (a) If λ in $\mathbb{R}$ is not an eigenvalue of $J_{d-1}(\tau)$, then it is an eigenvalue of $J_d(\tau)$ for some $\tau \in \mathbb{R}$. (b) Fix λ and find $\max\{\mu\{\lambda\} \mid \mu \text{ satisfies } (5.1)\}$.

References

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

N. I. Akhiezer and I. M. Glazman. *Theory of linear operators in Hilbert space.* Dover Publications, Inc., New York, 1993. ISBN 0-486-67748-6. Translated from the Russian and with a preface by Merlynd Nestell, Reprint of the 1961 and 1963 translations, Two volumes bound as one.

Thøger Bang. The theory of metric spaces applied to infinitely differentiable functions. *Math. Scand.*, 1:137–152, 1953. ISSN 0025-5521.

Thøger Bang. *Om quasi-analytiske Funktioner.* Thesis, University of Copenhagen, 1946.

Christian Berg. *Moment problems and orthogonal polynomials.* unpublished lecture notes, 2011.

Christian Berg. The multidimensional moment problem and semigroups. In *Moments in mathematics (San Antonio, Tex., 1987)*, volume 37 of *Proc. Sympos. Appl. Math.*, pages 110–124. Amer. Math. Soc., Providence, RI, 1987.

Christian Berg and Ryszard Szwarc. A determinant characterization of moment sequences with finitely many mass points. *Linear Multilinear Algebra*, 63(8): 1568–1576, 2015. ISSN 0308-1087. doi: 10.1080/03081087.2014.954516. URL `http://dx.doi.org/10.1080/03081087.2014.954516`.

Torsten Carleman. *Les fonctions quasi analytiques.* Gauthier-Villars, 1926.

Persi Diaconis. Application of the method of moments in probability and statistics. In *Moments in mathematics (San Antonio, Tex., 1987)*, volume 37 of *Proc. Sympos. Appl. Math.*, pages 125–142. Amer. Math. Soc., Providence, RI, 1987.

Jürgen Friedrich. A note on the two-dimensional moment problem. *Math. Nachr.*, 121: 285–286, 1985. ISSN 0025-584X.

Yitzhak Katznelson. *An introduction to harmonic analysis.* Cambridge Mathematical Library. Cambridge University Press, Cambridge, third edition, 2004. ISBN 0-521-83829-0; 0-521-54359-2.

A. G. Kostyučenko and B. S. Mityagin. The multi-dimensional problem of moments. *Soviet Math. Dokl.*, 1:415–419, 1960. ISSN 0197-6788.

M. G. Krein and A. A. Nudel′man. *The Markov moment problem and extremal problems.* American Mathematical Society, Providence, R.I., 1977. ISBN 0-8218-4500-4. Ideas and problems of P. L. Čebyšev and A. A. Markov and their further development, Translated from the Russian by D. Louvish, Translations of Mathematical Monographs, Vol. 50.

S. Mandelbrojt. Analytic functions and classes of infinitely differentiable functions. *Rice Inst. Pamphlet*, 29(1):142, 1942.

T. S. Motzkin. The arithmetic-geometric inequality. In *Inequalities (Proc. Sympos. Wright-Patterson Air Force Base, Ohio, 1965)*, pages 205–224. Academic Press, New York, 1967.

A. E. Nussbaum. Quasi-analytic vectors. *Ark. Mat.*, 6:179–191 (1965), 1965. ISSN 0004-2080.

G. Pólya. *Mathematics and plausible reasoning. Vol. II.* Princeton University Press, Princeton, NJ, second edition, 1990. ISBN 0-691-08006-2; 0-691-02510-X. Patterns of plausible inference.

Konrad Schmüdgen. The K-moment problem for compact semi-algebraic sets. *Math. Ann.*, 289(2):203–206, 1991. ISSN 0025-5831.

Arthur Wouk. Difference equations and J-matrices. *Duke Math. J.*, 20:141–159, 1953. ISSN 0012-7094.