Functional Analysis, the Moment Problem, and Semidefinite Programming

Patrick Oare

Resources

These notes are primarily based off Schmüdgen's textbook and additional resources on semidefinite programming.

- *Real and Functional Analysis* by Serge Lang.
- *Measure Theory* by Paul Halmos.

Introduction

The moment problem is the

1 Measure Theory and Function Spaces

We'll first sketch out some mathematical preliminaries (this includes basic measure theory). The space $\mathcal{X}$ denotes a locally compact Hausdorff space. For notation, denote

- $C(\mathcal{X})$ the space of continuous functions on $\mathcal{X}$,
- $C_c(\mathcal{X})$ denote the space of functions on $\mathcal{X}$ with compact support,
- $C_c(\mathcal{X}, \mathbb{R})$ the subspace of $C_c(\mathcal{X})$ which map $\mathcal{X} \to \mathbb{R}$,
- $C_c(\mathcal{X})_+$ the subspace of $C_c(\mathcal{X})$ consisting of non-negative functions, and
- $C_0(\mathcal{X})$ the $f \in C(\mathcal{X})$ which vanish at ∞: equivalently, the functions $f \in C(\mathcal{X})$ for which the set $\{x \in \mathcal{X} : |f(x)| \geq \epsilon\}$ is compact for all $\epsilon > 0$.

1.1 Signed and Complex Measures

We begin with some basics from measure theory. Namely, the extension of a measure to be complex-valued. We introduce measure theory in my previous note set on Measure Theory, but in that case, we only deal with the case of measures taking on positive (or infinite) values. In this case, a **positive measure** μ on a measurable space $(X,)$ is a map $\mu : \Sigma \to \overline{\mathbb{R}}_{\geq 0}$, where $\overline{R}_{\geq 0} = \mathbb{R}_{\geq 0} \cup \{\infty\}$ is the extended positive real numbers, which sends $\mu(\emptyset) = 0$ and is σ-additive (countably additive):

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i) \tag{1}$$

whenever $\{A_i\}_{i=1}^{\infty}$ is a countable collection of disjoint measurable sets $A_i \in \Sigma$.

The extension to a signed measure allows μ to take negative values, and to take infinite values, but only either $+\infty$ or $-\infty$.

Definition 1: Signed measure, complex measure

Let (X, Σ) be a measurable space. A **signed measure** is a map $\mu_{\mathbb{R}} : \Sigma \to \overline{\mathbb{R}}$ which has $\mu_{\mathbb{R}}(\emptyset) = 0$, takes on at most one of the values $+\infty$ or $-\infty$, and is σ-additive.

A **complex measure** is a map $\mu : \Sigma \to \mathbb{C}$ which has $\mu(\emptyset) = 0$ and is σ-additive. Note that complex measures do not take ∞ as a value.

The set of signed measures, and the set of complex measures, form a vector space over $\mathbb{R}$ and $\mathbb{C}$ respectively, with the definitions,

$$(\mu + \nu)(A) := \mu(A) + \nu(A) \qquad (c\mu)(A) := c\,\mu(A). \tag{2}$$

The reason a signed measure can only take at most one of the values $\{-\infty, +\infty\}$ is to allow $(+, \cdot)$ to give the set of signed measures a vector space structure. If only one is allowed, we can still define addition and scalar multiplication, because, if for example $\mu(A) = \infty$, then $(c\mu)(A) = \infty$, and $(\mu + \nu)(A) = \infty$ for any ν. If one allowed a signed measure to take on the values $+\infty$ and $-\infty$, then if $\mu(A) = +\infty$ and $\nu(A) = -\infty$, there is no consistent way to define the action of the measure $\mu + \nu$ on A. Likewise, we do not allow any value of ∞ to be taken by complex measures because we do not want to worry about adding different phases times infinity to one another. If you want to allow ∞ to be a value taken by a complex measure, one can do that, but then it is difficult to define a vector space structure on the space of measures.

Signed and complex measures generalize the notion of the integral of a non positive-definite function. Suppose μ is a positive measure and $f \geq 0$ is μ-integrable. It is common to define another positive measure by integrating f,

$$\nu(A) = \int_A d\mu\, f. \tag{3}$$

This has the interpretation of being the measure associated with density f, akin to how we integrate density functions in probability theory to yield another measure. Signed and complex measures allow us to use this definition when f can take on negative values, or complex values, assuming f it is relatively well-behaved (assuming $|f|$ is integrable). Indeed, we'll see with the Radon-Nikodym theorem that is is often the case that we can express an arbitrary measure ν in terms of another measure μ like this: we simply must ask that ν is absolutely continuous with respect to μ.

Definition 2: Regular Borel measure, Radon measure

Let X be a topological space. A (positive, signed, or complex) **Borel measure** μ is a (positive, signed, or complex) measure on X defined on the Borel σ-algebra $\mathcal{B}(X)$. We call μ a **regular Borel measure** if it satisfies **inner regularity**,

$$\mu(M) = \sup_{K \subseteq M,\ \text{K compact}} \mu(K), \tag{4}$$

and **outer regularity**,

$$\mu(M) = \sup_{U \supseteq M,\ \text{U open}} \mu(U), \tag{5}$$

for all measurable $M \in \mathcal{B}(X)$.

When X is a locally compact Hausdorff space, we denote it as $\mathcal{X}$. A **Radon measure** on $\mathcal{X}$ is a regular Borel measure such that $\mu(K) < \infty$ for each compact subset of $\mathcal{X}$. Typically, Radon measures are used when the domain is a locally compact Hausdorff space.

The set of Radon measures on $\mathcal{X}$ is denoted $M_+(\mathcal{X})$, the set of probability measures in $M_+(\mathcal{X})$ is denoted $M_+^1(\mathcal{X})$, and the set of finite measures in $M_+(\mathcal{X})$ is denoted $M_+^b(\mathcal{X})$, and the set of complex Radon measures is denoted $M(\mathcal{X})$. The set of Borel-measurable μ-integrable functions on $\mathcal{X}$ is denoted $\mathcal{L}^1(\mathcal{X}, \mu)$: a Borel function f is in $\mathcal{L}^1(\mathcal{X}, \mu)$ iff $\int d\mu\, |f| < \infty$.

Theorem 1: Lebesgue dominated convergence

Let $\mu \in M_+(\mathcal{X})$. Suppose that $g : \mathcal{X} \to \mathbb{R}_{\geq 0}$ is μ-integrable, and that f_n and f are complex Borel functions on $\mathcal{X}$ satisfying,

$$f_n \longrightarrow f \qquad\qquad |f_n(x)| \leq g(x) \tag{6}$$

μ-almost everywhere. Then the functions f_n and f are μ-integrable and converge in $\mathcal{L}^1(X, \mu)$,

$$\lim_{n\to\infty} \int d\mu\, |f_n - f| = 0 \qquad\qquad \lim_{n\to\infty} \int d\mu\, f_n = \int d\mu\, f. \tag{7}$$

In other words, a regular Borel measure is a measure on a topological space such that the measure every measurable set can be approximated from below by compact sets and above by open sets. When X is a locally compact Hausdorff space and is σ-compact (can be covered by

a countable number of compact sets), it suffices to just check that the measure is inner regular, and this implies outer regularity. This makes regular Borel and Radon measures work well with the topological structure on a space: most of the measures familiar to you on topological spaces (like the Lebesgue measure) are Radon measures. We'll be using Radon measures frequently in the context of the moment problem.

There are two important decompositions that are useful for signed measures and complex measures. Essentially, we can break a signed measure into a difference of two positive measures, and break a complex measure into a sum of two signed measures.

Theorem 2: Hahn-Jordan decomposition

Any signed measure μ on a measurable space (X, Σ) can be decomposed as,

$$\mu = \mu_1 - \mu_2, \tag{8}$$

where μ_1 and μ_2 are positive measures, and

$$X = E_1 \cup E_2 \qquad\qquad E_1 \cap E_2 = \emptyset, \tag{9}$$

where $E_1, E_2 \in \Sigma$ and $\mu_1(E_2) = 0 = \mu_2(E_1)$. The measures μ_1 and μ_2 are unique and called the *positive part* and *negative part* of μ, respectively, and E_1, E_2 are unique up to a $(\mu_1 + \mu_2)$-null set.

Definition 3: Support of a measure

Let μ be a Borel measure on a topological space $(X, \mathcal{B}(X))$. We say that μ is **supported** on the set $M \in \mathcal{B}(X)$ if

$$\mu(\mathcal{X} \setminus M) = 0. \tag{10}$$

The **support** of μ is the smallest closed subset $M \in \mathcal{B}(X)$ such that μ is supported on M, denoted $\operatorname{supp}\mu$. Two measures μ and ν are called **mutually singular** if they are supported on different subsets, $\operatorname{supp}\mu \cap \operatorname{supp}\nu = \emptyset$, and denoted $\mu \perp \nu$.

Note that $x \in \operatorname{supp}\mu$ iff every neighborhood $x \in U_x \in \mathcal{B}(\mathcal{X})$ has non-zero measure, $\mu(U_x) > 0$.

The Hahn Jordan decomposition explicitly decomposes a signed measure into the sum of a positive measure minus another positive measure which are mutually singular, $\mu_1 \perp \mu_2$. The set E_1 contains all the measurable subsets with positive μ-measure, and the set E_2 contains all the measurable subsets with negative μ-measure. The measure $\mu_1 + \mu_2$ is the **absolute value** of μ, denoted $|\mu|$, and is a positive measure. We an apply the Hahn Jordan decomposition to the real and complex parts of an arbitrary complex measure.

Corollary 1

Let μ be a complex Borel measure on $(X, \mathcal{B}(X))$. Then μ can be decomposed as,

$$\mu = (\mu_1 - \mu_2) + i(\mu_3 - \mu_4), \tag{11}$$

where $\mu_1 \perp \mu_2$, $\mu_3 \perp \mu_4$, and μ_i are positive measures.

Now we discuss relations between different measures.

Definition 4: Absolutely continuous

A measure ν is **absolutely continuous** with respect to a measure μ, denoted $\nu \ll \mu$, if every μ-null set A ($\mu(A) = 0$) is also a ν-null set ($\nu(A) = 0$). In other words, $\mu(A) = 0$ implies $\nu(A) = 0$.

Theorem 3: Radon-Nikodym

Let $\mu, \nu \in M(\mathcal{X})$ with μ σ-finite[a]. If $\nu \ll \mu$, then there is a function $h \in \mathcal{L}^1(\mathcal{X}, \mu)$ such that $d\nu(x) = h(x)\, d\mu(x)$. The function h is called the **Radon-Nikodym derivative** of ν with respect to μ and denoted $\frac{d\nu}{d\mu}$.

[a] $\mathcal{X}$ is a countable union of sets with finite μ-measure.

The Radon-Nikodym theorem makes formal what we were talking about before with integrating densities to define additional measures. It says that if an arbitrary measure ν behaves nicely with respect to μ, then ν can be interpreted as a change of variables on the original measure μ, where the Radon-Nikodym derivative gives the density which changes the measure. In this case, "behaves nicely" means that any atoms of ν are also atoms of μ: essentially it means that ν is a *rescaling* of μ up to a density, but if a set is μ-null, ν cannot give said set non-zero mass.

The easiest way to gain intuition for this is to consider probability measures with $\mu = \text{Leb}$ the Lebesgue measure on $\mathbb{R}$. If ν is a continuous probability measure, then ν does not give mass to Lebesgue-null sets, hence we can write

$$\nu(A) = \int_A d\nu(x) = \int_A d\mu(x)\, \frac{d\nu}{d\mu}(x) = \int_A dx \frac{d\nu}{d\mu}(x), \tag{12}$$

where $dx = d\text{Leb}(x)$ is the standard Lebesgue measure on $\mathbb{R}^n$. We see that the Radon-Nikodym derivative $d\nu/d\mu$ now has the interpretation of a **probability density function** on $\mathbb{R}^n$, and we should expect it to be normalized (if $\nu(\mathbb{R}^n) = 1$) and non-negative.

If ν is instead a mixed probability measure (a sum of discrete and continous distributions), we can no longer apply the Radon-Nikodym theorem to ν because ν gives mass to μ-null sets. Suppose that ν gives unit mass to discrete points $\{x_i\}_{i=1}^n \subseteq \mathbb{R}^n$. Then we can apply Radon-Nikodym to the measure,

$$\nu_{\text{cont}} := \nu - \sum_{i=1}^n \delta_{x_i}, \tag{13}$$

where δ_x denotes a Dirac mass at x. This expands ν as a sum of point masses plus a continuous distribution, where the density of the continuous distribution is the Radon-Nikodym derivative of ν_{cont}.

The final connection to make here is the connection between linear functionals and integration. Namely, for a given measure ν, the map,

$$C_c(\mathcal{X}, \mathbb{R}) \longrightarrow \mathbb{R} \qquad\qquad f \mapsto \int_{\mathcal{X}} d\nu\, f \tag{14}$$

defines a linear functional on $C_c(\mathcal{X}, \mathbb{R})$. Note that when $\nu \ll \mu$, we can equivalently think of this as the linear functional obtained by integrating f against $d\nu/d\mu$,

$$f \mapsto \int_{\mathcal{X}} d\mu\, \frac{d\nu}{d\mu}\, f. \tag{15}$$

Letting $g \in C_c(\mathcal{X}, \mathbb{R})$ be an arbitrary function with compact support, we can define the measure $d\nu(x) = g(x)\, d\mu(x)$, for which the functional $f \mapsto \int_{\mathcal{X}} f\, d\nu$ is exactly integration of f against the density g.

So, integration against an arbitrary measure ν (if $\nu \ll \mu$, equivalently against a density g) defines a linear functional on $C_c(\mathcal{X}, \mathbb{R})$. The converse of this question is to ask when a linear functional

$$L : C_c(\mathcal{X}, \mathbb{R}) \longrightarrow \mathbb{R}, \tag{16}$$

can be **represented** by a measure ν. That is, when can an arbitrary linear functional $L(f)$ be represented by integration against a measure (or integration against a density $d\mu/d\nu$),

$$L(f) \stackrel{?}{=} \int d\nu\, f. \tag{17}$$

This question is answered by the Riesz Representation Theorem, one of the cornerstones of functional analysis.

Theorem 4: Riesz Representation Theorem

For each linear functional $L : C_c(\mathcal{X}, \mathbb{R}) \longrightarrow \mathbb{R}$ such that $L(f) \geq 0$ for each $f \in C_c(\mathcal{X})_+$, there exists a unique positive Radon measure μ such that,

$$L(f) = \int_{\mathcal{X}} d\mu\, f, \tag{18}$$

for all $f \in C_c(\mathcal{X}, \mathbb{R})$.

1.2 Vague Convergence and Nets

Nets are an important concept in topology that allows us to generalize the notion of a sequence to other directed sets. Let us first explain what we mean by a "directed set".

Definition 5: Partially ordered set (poset)

Definition 6: Directed Set

A poset $(I, \leq)$ is a **directed set** if for each $a, b \in I$, there exists $c \in I$ such that,

$$a \leq c \qquad\qquad b \leq c. \tag{19}$$

1.3 Extensions of Functionals

One of the most fundamental results in functional analysis is the Hahn-Banach theorem, which allows one to extend linear functions from a subspace to the entire vector space, provided the functional is bounded by a sub-linear map.

Theorem 5: Hahn-Banach

Let X be a $\mathbb{R}$-vector space and $q : X \to \mathbb{R}$ a sub-linear[a] functional. If $M \leq X$ is a subspace and $f : M \to \mathbb{R}$ is a linear functional on M dominated by q ($f(x) \leq q(x)$ for all $x \in M$), then there exists an extension F of f to X that is also dominated by q on all of X. In other words,

$\exists F : X \to M$ such that,

$$F|M = f \qquad\qquad F(x) \leq q(x), \forall x \in X. \tag{20}$$

[a] By sub-linear, one means $q(x+y) \leq q(x) + q(y)$ and $q(\lambda x) = \lambda q(x)$ for $\lambda > 0$.

2 The Moment Problem

At its core, the moment problem is to determine a function given information about the moments of said function. That is, given the moments $(s_\alpha)_{\alpha\in\mathbb{N}_0}$,

$$s_\alpha = \int_{\mathcal{X}} dx\, f(x)\, x^\alpha, \tag{21}$$

how does one determine the original function f? With maximum generality, the space $\mathcal{X}$ will be a locally compact Hausdorff space, but for our purposes we will often focus on the case when $\mathcal{X} = \mathbb{R}$ (the Hamburger moment problem), when $\mathcal{X} = \mathbb{R}_{\geq 0}$ (the Stieltjes moment problem), or when $\mathcal{X} = [0, 1]$ (the Hausdorff moment problem). In theory one often considers the moment problem when all the moments of f are known, but in practice one is often more interested in the **truncated moment problem** in which only a finite number of moments of f are known.

The problem can be phrased measure-theoretically as well, which will be the language we use in these notes. In this case, given a collection of moments (s_α), we wish to determine which Radon measures ν satisfy,

$$s_\alpha = \int_{\mathcal{X}} d\nu(x)\, x^\alpha. \tag{22}$$

This is the same problem as above, just writing it in terms of the measure ν instead of the density $g(x)$, with $d\nu(x) = g(x)dx$.

This can be abstracted even further, and will be formulation of the moment problem that we study. Given a linear functional L on a vector space E of continuous functions on a locally compact Hausdorff space $\mathcal{X}$ and a closed subset $K \subseteq \mathcal{X}$, when does there exist a (positive) Radon measure μ supported on K such that,

$$L(f) = \int_{\mathcal{X}} d\mu(x)\, f(x), \tag{23}$$

for each $f \in E$? Such a functional is called a **K-moment functional**, and when $K = E$ is simply called a **moment functional**. Note that the previous definition of the moment problem is equivalent to this formulation, for the specific subspace,

$$E = \text{span}\{1, x, x^2, x^3, ...\} = \mathbb{R}[x], \tag{24}$$

where $\mathbb{R}[x]$ denotes the polynomial ring over $\mathbb{R}$ in the indeterminate x.

For some notation, given a subset $C \subseteq L$, a functional L is called **C-positive** if $L(f) \geq 0$ for each $f \in C$. We let,

$$E_+ := \{f \in E : f(x) \geq 0\, \forall x \in \mathcal{X}\}. \tag{25}$$

Given a positive Radon measure $\mu \in M_+(\mathcal{X})$, such that $E \subseteq \mathcal{L}^1(\mathcal{X}, \mathbb{R})$ is integrable, we define the E_+-positive functional L^μ by integration against μ,

$$L^\mu(f) := \int_{\mathcal{X}} d\mu(x)\, f(x). \tag{26}$$

Any functional L which has a form $L = L^\mu$ is called a **moment functional**, and μ is a **representing measure** of L. The **set of representing measures of** L is denoted,

$$\mathcal{M}_L := \{\mu \in M_+(\mathcal{X}) : L = L^\mu\}. \tag{27}$$

If $|\mathcal{M}_L| = 1$ (L has a unique representing measure), then L is called **determinate**. If the measures are supported on $K \subseteq \mathcal{X}$, we use an additional K subscript on the previously defined notation.

Given a sequence of moments $(s_\alpha)_{\alpha\in\mathbb{N}_0}$, the **Riesz functional** L_s on $\mathbb{R}[x]$ is defined by setting,

$$L_s : \begin{cases} \mathbb{R}[x] \longrightarrow \mathbb{R} \\ x^n \longmapsto s_n, \forall n \in \mathbb{N}_0 \end{cases}, \tag{28}$$

and extending L_s to arbitrary $p \in \mathbb{R}[x]$ by linearity. The moment problem is thus: what are the conditions on the sequence (s_α) so that $L_s = L^\mu$ for some Radon measure $\mu \in M_+(\mathcal{X})$?

The next part of this section concerns the existence of representing measures. Many of the existence results in this area stem from this and similar theorems, developed by Choquet. We first explore the idea of an *adapted subspace* of $C(\mathcal{X}, \mathbb{R})$.

Definition 7: Adapted subspace

A space $E \leq C(\mathcal{X}, \mathbb{R})$ is called **adapted** if the following three conditions hold:

1. $E = E_+ - E_+$.
2. For each $x \in \mathcal{X}$, $\exists f \in E_+$ such that $f(x) > 0$.
3. For each $f \in E_+$, there is $g \in E_+$ such that g dominates f.

My heuristic definition of adapted subspaces is that they act like spaces of bump forms. The first condition means that one should define an adapted space with a basis of purely positive or negative functions (while this might not be true in all cases, it seems to roughly hold). For example with $\mathcal{X} = \mathbb{R}$, the space $\text{span}\{x \mapsto x^3\}$ fails condition (1) since $E_+ = \{0\}$. On the other hand, $E = \text{span}\{x \mapsto x^2\}$ satisfies condition 1 (although not condition (3)). The second condition just means that we can separate x from zero using a function in E: again, here my intuition is that of a bump form centered at x, but it can be any function in E_+ with $f(x) > 0$. The last condition is perhaps the most mysterious.

An example of an adapted space is,

$$E = C_c(\mathbb{R}, \mathbb{R}), \tag{29}$$

that is, the space of compactly supported functions $\mathbb{R} \to \mathbb{R}$. The first condition is satisfied because any $f \in E$ can be decomposed as $f = f_+ - f_-$, where $f_+ := f \vee 0$ and $f_- = (-f) \vee 0$ are both positive functions with compact support. The last condition should be trivially satisfied as well because all the functions have compact support. A more interesting case is,

$$E = C_0(\mathbb{R}, \mathbb{R}), \tag{30}$$

the space of functions $\mathbb{R} \to \mathbb{R}$ which vanish at ∞. This likewise satisfies the first two conditions to be an adapted space, but the third condition is now non-trivial.

2.1 Existence and Determinance of Solutions

The moment problem for $(s_\alpha)_{\alpha\in\mathbb{N}_0}$ is called **determinate** if there is a unique solution, i.e. if $L_s = L^\mu$ for a unique representing measure $\mu \in M_+(\mathcal{X})$. In this section we explore the existence of solutions and the determinacy of said solutions.

A moment sequence (s_α) is **positive semi-definite (PSD)** if for all $\xi_1, ..., \xi_n \in \mathbb{R}$ and $n \in \mathbb{N}$, we have,

$$\sum_{k,\ell=0}^{n} s_{k+\ell}\xi_k\xi_\ell \geq 0 \tag{31}$$

(the sequence is positive definite if $\geq$ is strict). We denote the set of all PSD moment sequences as $\mathcal{P}(\mathbb{N}_0)$. The **shift operator** E on a sequence is defined as,

$$(Es)_n := s_{n+1}. \tag{32}$$

The most important object to determine if a moment problem has a solution is the **Hankel matrix** $H(s)$ and its determinant $D(s)$.

Definition 8: Hankel matrix, Hankel determinant

The Hankel matrix is an $(n+1)\times(n+1)$ matrix of moments which is constant on the diagonal,

$$H_n(s) := \begin{pmatrix} s_0 & s_1 & s_2 & \cdots & s_n \\ s_1 & s_2 & s_3 & \cdots & s_{n+1} \\ s_2 & s_3 & s_4 & \cdots & s_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & s_{n+2} & \cdots & s_{2n} \end{pmatrix} \qquad D(s) := \det H(s). \tag{33}$$

In other words, $H_n(s)_{k,\ell} = s_{k+\ell}$. A sequence s is PSD, i.e. $s \in \mathcal{P}(\mathbb{N}_0)$, if and only if all its Hankel matrices are PSD. This has a number of immediate corollaries, which are summarized in **Hamburger's theorem**.

Theorem 6: Hamburger

Let $(s_n)_{n\in\mathbb{N}_0}$ be a real sequence. Then TFAE:

1. (s_n) is a Hamburger moment sequence: there exists a Radon measure $\mu \in M_+(\mathbb{R})$ such that for all $n \in \mathbb{N}_0$, $x^n \in \mathcal{L}^1(\mathbb{R}, \mu)$, and,
$$s_n = \int_{\mathbb{R}} d\mu(x)\, x^n. \tag{34}$$
2. $(s_n) \in \mathcal{P}(\mathbb{N}_0)$ is PSD.
3. For each $n \in \mathbb{N}_0$, $H_n(s) \succeq 0$.
4. L_s is a positive functional on $\mathbb{R}[x]$, i.e., for all $p \in \mathbb{R}[x]$,
$$L_s(p^2) \geq 0. \tag{35}$$

3 Orthogonal Polynomials and the Moment Problem

Recall the Riesz functional $L_s \in \mathbb{C}[x]^*$ is defined by sending each polynomial to its corresponding moment,

$$L_s : x^n \longmapsto s_n. \tag{36}$$

If L_s is a moment functional on $\mathcal{X}$ represented by a Radon measure $\mu \in M_+(\mathcal{X})$, then this has the interpretation of being integration against μ,

$$L_s(f) = \int_{\mathcal{X}} d\mu\, f. \tag{37}$$

This functional will be a topic of study in its own right, and in this section we will study how it induces an inner product on $\mathbb{R}[x]$ (or $\mathbb{C}[x]$) and study the orthogonal polynomials with respect to this inner product. For this chapter, assume $(s_n)_{n\in\mathbb{N}_0}$ is a positive-definite moment sequence, so it has a solution and is not determinate. We will work over $\mathbb{C}[x]$.

For $p, q \in \mathbb{C}[x]$, we define a pairing,

$$\langle p, q\rangle_s := L_s(p\overline{q}). \tag{38}$$

Note that by definition this is skew-symmetric and linear in its first argument, and because (s_n) is PSD, this is non-negative definite,

$$\langle p, p\rangle_s = L_s(|p|^2) = L_s\left(\sum_{i,j} p_i \overline{p}_j x^{i+j}\right) = \sum_{i,j} s_{i+j} p_i \overline{p}_j > 0, \tag{39}$$

with $\langle p, p\rangle_s = 0$ if and only if $p = 0$. This hence defines an inner product, and **we can view $(\mathbb{C}[x], \langle\cdot,\cdot\rangle_s)$ as an inner product space.** Note, though, that this is not necessarily a Hilbert space, as we do not know if it is complete under $||\cdot||_s$! We will later consider the Hilbert space completion of this inner product space.

By applying Gram-Schmidt to the basis $\{1, x, x^2, ...\}$ of $\mathbb{C}[x]$, one can form bases for $\mathbb{C}[x]$ consisting of orthogonal polynomials. We will study two particular bases in this section: the sequence $(p_n)_{n\in\mathbb{N}_0}$ of orthonormal polynomials, and the sequence $(P_n)_{n\in\mathbb{N}_0}$ of monic, orthogonal polynomials. The sequence $(p_n)_{n\in\mathbb{N}_0}$ are also called the **orthogonal polynomials of the first kind**. Both are uniquely defined once the moment sequence (s_n) is given.

Next, we construct a few formulas for (p_n) in terms of the determinant of the Hankel matrix,

$$D_n = \det H_n(s) = \det \begin{pmatrix} s_0 & s_1 & s_2 & \cdots & s_n \\ s_1 & s_2 & s_3 & \cdots & s_{n+1} \\ s_2 & s_3 & s_4 & \cdots & s_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_n & s_{n+1} & s_{n+2} & \cdots & s_{2n} \end{pmatrix}. \tag{40}$$

Importantly, we will also construct a 3-term recurrence relation that can be used to define the orthogonal polynomials.

The simplest formula for (p_n) is to express it in terms of a modified Hankel determinant, with

the bottom row being replaced by monomials,

$$p_n(x) = \frac{1}{\sqrt{D_n D_{n-1}}} \det \begin{pmatrix} s_0 & s_1 & s_2 & ... & s_n \\ s_1 & s_2 & s_3 & ... & s_{n+1} \\ s_2 & s_3 & s_4 & ... & s_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_n & s_{n+1} & ... & s_{2n-1} \\ 1 & x & x^2 & ... & x^n \end{pmatrix}, \qquad p_0 = s_0^{-1/2}. \tag{41}$$

The leading coefficient of p_n is,

$$\langle x^n, p_n \rangle_s = \sqrt{\frac{D_n}{D_{n-1}}}, \tag{42}$$

and for $k < n$, we have orthogonality

$$\langle x^k, p_n \rangle_s = 0. \tag{43}$$

Note that because we are using the $\langle \cdot, \cdot \rangle_s$ brackets, orthogonality does not imply that p_n is a monomial of degree n: indeed, it is still a linear combination of $\{1, x, x^2, ..., x^n\}$.

In order to prove this claim, we show that Eq. (41) defines a sequence $(p_n)_{n \in \mathbb{N}_0}$ which satisfies the orthogonality condition (Eq. (43)) with a given leading coefficient (Eq. (42)). This then implies that the (p_n) defined in Eq. (41) are indeed a valid equation for the orthonormal polynomials, since they are uniquely defined and this definition is orthonormal with respect to $\langle \cdot, \cdot \rangle_s$. So, we show that Eq. (41) satisfies the two conditions. Note that,

$$\langle x^k, p_n \rangle = L_s(x^k \overline{p}_n(x)), \tag{44}$$

and because the determinant is row-linear, we can move the x^k factor inside the determinant defining p_n,

$$x^k p_n(x) = \frac{1}{\sqrt{D_n D_{n-1}}} \det \begin{pmatrix} s_0 & s_1 & s_2 & ... & s_n \\ s_1 & s_2 & s_3 & ... & s_{n+1} \\ s_2 & s_3 & s_4 & ... & s_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_n & s_{n+1} & ... & s_{2n-1} \\ x^k & x^{k+1} & x^{k+2} & ... & x^{k+n} \end{pmatrix} \tag{45}$$

Evaluating $\langle x^k, p_n \rangle_s$ then simply sends each x^j to s_j because of linearity of the integral, hence,

$$\langle x^k, p_n \rangle = \frac{1}{\sqrt{D_n D_{n-1}}} \det \begin{pmatrix} s_0 & s_1 & s_2 & ... & s_n \\ s_1 & s_2 & s_3 & ... & s_{n+1} \\ s_2 & s_3 & s_4 & ... & s_{n+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ s_{n-1} & s_n & s_{n+1} & ... & s_{2n-1} \\ s_k & s_{k+1} & s_{k+2} & ... & s_{k+n} \end{pmatrix}, \tag{46}$$

which vanishes for $k < n$ because it has repeated rows. For $k = n$, this is exactly the Hankel determinant,

$$\langle x^n, p_n \rangle_s = \frac{1}{\sqrt{D_n D_{n-1}}} D_n = \sqrt{\frac{D_n}{D_{n-1}}}, \tag{47}$$

as claimed.

3.1 The Jacobi operator

4 The Truncated Moment Problem

Many of the ideas used in studying the full moment problem also have analogs when the number of moments known is finite. For this chapter, suppose we know moments $(s_j)_{j=0}^m$. We can construct a Hankel matrix $H_n(s)$ as previously for any $n \in \mathbb{N}_0$ with $2n \leq m$, and the Riesz functional L_s is now defined on the finite-dimensional algebra of degree $\leq m$ polynomials, $\mathbb{C}[x]_m := \{p \in \mathbb{C}[x] : \deg p \leq m\}$,

$$L_s : \mathbb{C}[x]_m \longrightarrow \mathbb{C} \qquad\qquad L_s(x^j) \longmapsto s_j. \tag{48}$$

5 Semidefinite Programming (SDP)

Semidefinite programming is a method that can be used to study the moment problem. In general, it refers to solving the optimization problem,

$$\min_{X \in \text{Herm}(N)} C \cdot X \text{ s.t. } A_i \cdot X - b_i = 0, \ \forall i \in \{1, ..., m\} \text{ and } X \succeq 0. \tag{49}$$

Here $\text{Herm}(N)$ is the space of Hermitian $N \times N$ matrices, the matrix $C, A_i \in \text{Herm}(N)$ and b_i are arbitrary vectors. The inner product $C \cdot X$ is the Hermitian inner product, which will be expanded on in the next section.

5.1 Symmetric (Hermitian) Matrices

Hermitian matrices play an important role in understanding some of the features of SDP. The space $\text{Herm}(N)$ comes equipped with an inner product,

$$X \cdot Y := \sum_{i,j} X_{ij} Y_{ij} = \text{Tr}\, X^\dagger Y. \tag{50}$$

This Hermitian inner product induces the Frobenius norm,

$$||X||^2 = ||X||_F^2 = \sum_{i,j} |X_{ij}|^2. \tag{51}$$

Note that $\text{Herm}(N)$ is indeed a vector space over $\mathbb{C}$, as it is closed under $+$ and scalar multiplication.

Recall that the **spectral theorem** says that a Hermitian matrix X is unitarily diagonalizable with real eigenvalues,

$$X = U \Lambda U^\dagger, \tag{52}$$

where if $X u_i = \lambda_i u_i$ furnishes the eigensystem of X, then $U = \text{col}\, u_i$ is the matrix of orthonormal eigenvectors and $\Lambda = \text{diag}\{\lambda_1, ..., \lambda_N\}$ is the diagonal matrix of eigenvalues. Note that this similarity implies the Frobenius norm and the 2-norm of $X \in \text{Herm}(N)$ can be written in terms of the eigenvalues of X; in particular,

$$||X||_F = \sqrt{\sum_{i=1}^{N} \lambda_i^2} \qquad\qquad ||X||_2 = \max_{i=1}^{N} |\lambda_i|. \tag{53}$$

The eigenvalues of X are given by its characteristic polynomial. If $X(\epsilon)$ depends on some free parameter ϵ and is smooth with respect to ϵ, this implies that the eigenvalues of $X(\epsilon)$ are smooth with respect to ϵ as well, i.e. $\lambda_i = \lambda_i(\epsilon)$ is smooth.

It is often of interest to characterize the ordering of a symmetric matrix A, especially when said matrix depends on external parameters $A(x)$. The **Rayleigh-Ritz** and **Courant-Fisher** theorems help to characterize these eigenvalues. Denote the eigenvalue ordering of A by,

$$\lambda_1(A) \geq \lambda_2(A) \geq ... \geq \lambda_N(A). \tag{54}$$

Theorem 7: Rayleigh-Ritz

The extremal eigenvalues of A are given by,

$$\lambda_1(A) = \max_{x \in \mathbb{R}^N \setminus \{0\}} \frac{x^\dagger A x}{x^\dagger x} \qquad \lambda_N(A) = \min_{x \in \mathbb{R}^N \setminus \{0\}} \frac{x^\dagger A x}{x^\dagger x} \tag{55}$$

Theorem 8: Courant-Fisher

For $1 \leq k \leq n$, the k^{th} largest eigenvalue of A is,

$$\lambda_k(A) = \max_{S \leq \mathbb{R}^N, \dim S = k} \min_{x \in S \setminus \{0\}} \frac{x^\dagger A x}{x^\dagger x}. \tag{56}$$

In other words, the k^{th} largest eigenvalue can be found by projecting A onto a rank k subspace and finding the minimal eigenvalue on that subspace, then varying this subspace and maximizing this minimal eigenvalue. Both of these theorems vary the Rayleigh quotient $x^\dagger A x / ||x||_2^2$ in order to determine an estimate for the eigenvalues of A.

6 Semidefinite Programming and the Moment Problem